Title: Repetitive sequences in malaria parasite proteins

Authors: Heledd M. Davies[1,2*], Stephanie D. Nofal[1,3*], Emilia J. McLaughlin[1], Andrew R. Osborne[1]

Summary: Repetitive sequences are widespread in *Plasmodium falciparum* proteins; the importance of these sequences is increasingly apparent.

Keywords: Malaria, *Plasmodium falciparum*, Protein repeats, low-complexity, host-pathogen interaction, protein evolution

ABSTRACT

Five species of parasite cause malaria in humans with the most severe disease caused by *Plasmodium falciparum*. Many of the proteins encoded in the *P. falciparum* genome are unusually enriched in repetitive low complexity sequences containing a limited repertoire of amino acids. These repetitive sequences expand and contract dynamically and are among the most rapidly changing sequences in the genome. The simplest repetitive sequences consist of single amino acid repeats such as poly-asparagine tracts which are found in approximately 25% of *P. falciparum* proteins. More complex repeats of two or more amino acids are also common in diverse parasite protein families. There is no universal explanation for the occurrence of repetitive sequences and it is possible that many confer no function to the encoded protein and no selective advantage or disadvantage to the parasite. However, there are increasing numbers of examples where repetitive sequences are important for parasite protein function. We discuss the diverse roles of low complexity repetitive sequences throughout the parasite life cycle, from mediating protein-protein interactions to enabling the parasite to evade the host immune system.

INTRODUCTION

Repetitive nucleotide sequences expand and contract via strand slippage during DNA replication and unequal crossing over during meiosis (Fig. 1)(Gemayel et al., 2010). Within the genome of the *P. falciparum* strain 3D7, repetitive sequences are found on average once every 467 nucleotides (Tan et al., 2010). These repeat encoding sequences tend to be highly dynamic; the rate of insertion and deletion mutations in genomes of parasites propagated in the asexual blood stage is approximately 10 fold greater than the rate of single base pair substitution (Hamilton et al., 2016) and within coding sequences, the majority of insertion and deletion mutations occur in short tandemly repeated sequences (Hamilton et al., 2016, Miles et al., 2016). The variability of these sequences is also highlighted by the observation that 9% of genes contain repetitive sequences that vary in length between *P. falciparum* strains 3D7, Dd2, and HB3 (Tan et al., 2010). Perfect repeats are predicted to have expanded more recently than degenerate repeats and are particularly common in *Plasmodium* species (Mendes et al., 2013).

Many short tandemly repeated sequences within *P. falciparum* proteins are predicted to be intrinsically disordered (Feng et al., 2006, Guy et al., 2015). In isolation, such sequences are likely dynamic and do not fold into stably folded structures but exist as an ensemble of conformations (Oldfield and Dunker, 2014). In many cases, these sequences may not have a function *per se* but simply form disordered loops or termini protruding from folded domains. However, repeat sequences can also significantly change the properties of proteins and over evolutionary time expansion and contraction of repeats may fine-tune protein activity or even generate sequences *de novo* that can confer novel function. Indeed, low complexity

sequences, which include many repetitive sequences, are highly diverse. Whilst many are evolving neutrally others are likely under selective pressure (Battistuzzi et al., 2016, Zilversmit et al., 2010).

Diverse functions for repetitive protein sequences in diverse organisms

Functional repetitive proteins from many species have been extensively characterized and illustrate many of the concepts that will also apply to this class of polypeptides from *Plasmodium*. Stretches of single amino acid repeats are often prone to self-association, leading to the formation of protein aggregates or fibres (Oma et al., 2004). The detrimental effects of poly-glutamine and poly-alanine sequence aggregation are well-understood as they are associated with many human diseases (Gatchel and Zoghbi, 2005, Shoubridge et al., 2007). Expansion of these repeat-encoding sequences increases their aggregation propensity and contributes to disease onset. Expansion of single amino acid repeats is also important in non-pathological contexts; the activity of many yeast transcription factors can be fine-tuned over evolutionary time by expansion or contraction of poly-glutamine sequences (Gemayel et al., 2015). Artificial homo-polymeric sequences composed of different amino acids can direct proteins to different cellular compartments in mammalian cells (Oma et al., 2004) and poly-histidine repeats in several human DNA-binding proteins can confer localization to nuclear speckles (Salichs et al., 2009). *P. falciparum* contains a striking number of asparagine-rich tracts; similar sequences have been shown to form fibres in yeast prion proteins which are believed to regulate protein activity and allow the cell to respond to environmental stimuli (Chernova et al., 2014, Halfmann et al., 2011, Halfmann et al., 2010, Peters and Huang, 2007).

Notably, there are several examples in which repeat length can influence the phenotype of an organism. For example, experimentally changing the number of repeats in the *S. cerevisiae* cell surface proteins Flo1p and Flo11p proportionally changes the adhesion and hydrophobicity of the cells, respectively (Fidalgo et al., 2006, Verstrepen et al., 2005). In the case of Flo1p, a central repeat sequence forms a linker that allows an N-terminal adhesion domain to extend away from the cell surface; expansion of this sequence increases the efficiency of cell adherence. The repetitive sequence in Flo11p is hydrophobic and repeat expansion therefore increases the overall hydrophobicity of the yeast cell surface, allowing cells expressing Flo11p with longer repeats to float on the surface of a liquid medium. Repeat sequences can also contribute to altered phenotypes in more complex organisms; changes in repeat sequences in morphogen proteins are likely responsible for some of the different morphological features that have been selected for in domesticated dogs (Fondon and Garner, 2004, Fondon and Garner, 2007).

These examples establish several key concepts: firstly, repeat sequences can be important for protein function *per se*; secondly, expansion or contraction of repetitive sequences can cause partial or complete loss or gain of a protein function; thirdly, changes in protein activity arising from changes to repeat length can cause phenotypic changes, the magnitude of which can be proportional to the repeat length; and finally, expansion and contraction of repeat sequences can potentially facilitate adaption of an organism over evolutionary time.

Protein interactions and structures of repetitive protein sequences

Although relatively few repetitive sequences from *Plasmodium* proteins have been extensively characterized many of the applicable concepts relating to their structure and function have been well established in the context of repetitive proteins from other organisms

(Andrade et al., 2001, Kajava, 2012). In isolation many repetitive sequences are intrinsically disordered but upon interaction with a binding partner will assume a structured conformation. The structural plasticity and repeating modular architecture of repetitive low complexity sequences results in binding interactions that can differ from those typically described for globular proteins.

Repetitive protein sequences can potentially recruit multiple copies of an interacting partner. For example, multiple copies of transcriptional terminator proteins can interact cooperatively with neighbouring repeats in the C-terminal domain (CTD) of yeast RNA polymerase II (Lunde et al., 2010). Interaction with multiple copies of a binding partner can also increase interaction avidity. For example, the *Staphylococcus aureus* protein FnBPA contains short repeated motifs that interact with tandemly repeated β-sheet domains in human fibronectin. Although in isolation FnBPA repeats are intrinsically disordered, each repeat within a single FnBPA protein interacts with up to four consecutive fibronectin domains by contributing an additional β-strand to each domain (Bingham et al., 2008). As FnBPA contains multiple fibronectin-binding repeats it can interact with multiple fibronectin proteins which is important for adhesion to and invasion of endothelial cells (Edwards et al., 2010).

Alternatively, multiple repeats within a sequence can simultaneously interact with a single non-repetitive protein as illustrated by the interaction of the RNA polymerase II CTD with the mediator complex; here the repeats interact with four structurally non-identical binding sites with each identical repeat assuming a different conformation (Robinson et al., 2012). In addition, the CTD of RNA polymerase II interacts with many other proteins in an orchestrated manner that is regulated by its post-translational modification. This illustrates the key paradigm that repetitive sequences can function as regulated protein interaction hubs (Cumberworth et al., 2013). In this context the role of the *P. falciparum* RNA polymerase II CTD is discussed in a later section.

Although many repetitive sequences acquire structure only upon binding to a structured binding protein some form structured domains in isolation. For example, the repeated three amino acid motifs in collagen assemble to form a right-handed triple helix (Shoulders and Raines, 2009). Recent data also indicate that some repetitive sequences have a propensity to self-associate and undergo phase separation and thus partition into membraneless organelles within the cell (Kato et al., 2012). Although in most cases the functions and interactions of repetitive sequences in *Plasmodium* proteins are not well understood it is likely that these same principles will be relevant to many proteins that are expressed throughout the parasite lifecycle.

THE MALARIA PARASITE LIFECYCLE

Repetitive proteins are expressed throughout the malaria parasite lifecycle during which the parasite undergoes several differentiation steps within both insect and vertebrate hosts (Fig. 2). The repetitive proteins discussed in this review, their repeat sequences, the variability of repeats between different *P. falciparum strains*, and time of expression during the lifecycle are summarized in table 1.

An infection is initiated when haploid sporozoites are injected into the human host as a female *Anopheles* mosquito takes a blood meal. Having migrated via the bloodstream to the liver, sporozoites invade hepatocytes and multiply into thousands of haploid merozoite forms before being released back into the blood (reviewed in (Lindner et al., 2012)). The released merozoites invade erythrocytes and replicate intracellularly (reviewed in (Cowman et al., 2016)). After replication within the erythrocyte, daughter merozoites egress and after a brief

extracellular period invade new erythrocytes, undergoing further rounds of replication, egress and reinvasion. During the blood stage a fraction of the parasite population becomes committed to differentiate into gametocytes, the sexual stage of the parasite lifecycle. Gametocytes can be taken up by a mosquito during a blood meal (reviewed in (Bennink et al., 2016)). Within the mosquito midgut male gametocytes undergo exflagellation to form motile microgametes and female gametocytes form macrogametes. Gamete fusion leads to formation of a diploid zygote that differentiates into a motile ookinete which penetrates the mosquito midgut and forms an oocyst. Haploid sporozoites develop within the oocyst and are eventually released into the mosquito hemolymph before travelling to the salivary glands from where they can be injected into the next host (reviewed in (Aly et al., 2009)). After gamete fusion the parasite is briefly diploid but following a meiotic division the parasite returns to a haploid state and replicates mitotically for the remainder of the lifecycle. It is likely that expansion and contraction of gene sequences encoding repetitive proteins occurs in both meiotic and mitotic phases of the lifecycle.

FUNCTIONS OF REPEAT PROTEINS ON THE PARASITE SURFACE, WITHIN THE PARASITE AND IN THE INTERACTION WITH HOST CELLS

Whilst many proteins expressed throughout the lifecycle contain repetitive sequences the roles of these sequences are in most cases poorly understood. In the following sections we will discuss how repetitive proteins interact with the host immune system as well as specific proteins and groups of proteins where the function or importance of repeat sequences have been established.

Repetitive *Plasmodium* proteins and the immune response

Within the human host many repetitive parasite proteins are recognised by the immune system. 79% of tandem repeats in parasite proteins fall within protein sequences predicted to be intrinsically disordered (Guy et al., 2015). These sequences, in comparison to those from globular folded proteins, have a biased amino acid composition which results in fewer predicted MHCI and II peptides but a greater number of predicted B-cell epitopes (Guy et al., 2015, Hughes, 2004, Verra and Hughes, 1999). Notably, antibody interactions with disordered epitopes are predicted to be more sensitive to epitope sequence changes than interactions with structured epitopes. Given the propensity of disordered sequences to tolerate sequence polymorphism, this may allow effective evasion of antibody responses against these sequences (MacRaild et al., 2016).

Although sera from individuals living in endemic regions contain high levels of antibodies against repetitive proteins it is unclear whether these confer protection against disease. It has been speculated that repetitive proteins can even be detrimental to an effective immune response by crosslinking B-cell receptors and inducing an inferior T-cell independent immune response (Schofield, 1991). Although the antibody responses to most *Plasmodium* repetitive antigens have not been characterized, in mice there is a T-cell independent B-cell response to a repetitive sequence in the circumsporozoite protein (discussed later) but the bulk of the B-cell response is T-cell dependent (Fisher et al., 2017). It has also been proposed that repetitive antigens can generate an ‘antigenic smokescreen’ in which numerous cross-

reactive repetitive antigens interfere with the affinity maturation of antibodies against protective antigens (Kemp et al., 1987).

Nonetheless, many repetitive sequences are recognised by the immune system, and in particular those located on the parasite surface potentially represent important vaccine components and targets of host immunity. For the majority of the forty eight hour asexual replication cycle in the blood stage the parasite resides within the host erythrocyte. However, after egress merozoites are extracellular for up to several minutes before they invade new erythrocytes. During this period many repetitive merozoite surface proteins including MSP1, MSP2, GLURP, and SERA5 (SERA5 is a parasitophorous vacuole protein important for egress (Collins et al., 2017) but remains associated with the merozoite surface (Li et al., 2002)) and others are exposed to the host immune system (Beeson et al., 2016). SERA5 antibodies isolated from serum of Ugandan adults predominantly recognise the repetitive sequences within the protein and also show parasite growth inhibition *in vitro* in the presence of monocytes (Yagi et al., 2014). Similarly, antibodies that recognise the repetitive sequences in MSP1 and GLURP also show parasite growth inhibition (Galamo et al., 2009, Theisen et al., 1998). Alleles of the repetitive sequence in MSP1 appear to be under balancing selection which is consistent with this sequence being under immune selection (Conway et al., 2000). The repeating motifs of both MSP1 and MSP2 vary between isolates (Table 1), suggesting that there is not a strong functional constraint on these sequences and that they might easily change in response to immune pressure.

Circumsporozoite protein (CSP) is the major protein on the surface of the sporozoite which is also extracellular in the human host (Sinnis and Nardin, 2002). Both T-cell and antibody responses are elicited against CSP during infection. The RTS,S vaccine, which is currently the most successful malaria vaccine, provides partial protection against disease (Hoffman et al., 2015, Rts, 2015) and is based on a fragment of the circumsporozoite protein comprising tandemly repeated NANP motifs and a C-terminal domain. The central repetitive sequence contains immunodominant antibody epitopes (Zavala et al., 1985). Significantly, the repeat sequence in CSP is variable between parasite species. The repeating sequence motif in *P. berghei* CSP differs from those found in the *P. falciparum* and *P. vivax* proteins but despite this the repeat sequence from *P. berghei* CSP can be replaced with the repeat sequence from the *P. vivax* or *P. falciparum* protein without compromising function (Espinosa et al., 2013, Persson et al., 2002). Also, the number of repeats varies from 37 to 44 in different strains of *P. falciparum* (Neafsey et al., 2015). Although variation in the number of repeat units in CSP does not significantly affect vaccine efficacy (Neafsey et al., 2015) significant sequence variation can clearly be tolerated in the repeat sequences and their mutation may allow the parasite to evade B-cell responses without disrupting the function of CSP.

Circumsporozoite protein – a repetitive sequence essential for sporozoite development

Circumsporozoite protein, which is essential for parasite survival (Ferguson et al., 2014), is linked via a C-terminal glycophosphatidylinositol (GPI) anchor to the sporozoite plasma membrane. It consists of an N-terminal domain that binds heparin sulphate proteoglycans, a repeating central sequence, and a C-terminal Thrombospondin repeat domain (Sinnis and Nardin, 2002). Deletion of the central repeating sequence in the *P. berghei* CSP gene leads to a defect in sporozoite development in the mosquito. Although oocyst formation appears to be normal, the development of sporozoites is defective and sporozoites fail to appear in either the hemolymph or salivary glands of the mosquito (Ferguson et al., 2014).

While many repetitive sequences are predicted to be entirely disordered, biophysical analyses are consistent with the CSP repeats adopting an elongated but flexible cylinder-like structure that is about 18nm in length and 1.5nm in diameter and composed of β-turn structures (Plassmeyer et al., 2009). The model is based on the propensity of short NANP peptides to adopt a β-turn conformation (Ghasparian et al., 2006, Nanzer et al., 1997). NMR and circular dichroism studies suggest that CSP repeat peptides are in equilibrium between folded and extended states (Dyson et al., 1990, Fisher et al., 2017). Consistent with this, experiments examining the force required to unfold the CSP repeat region support a model in which CSP repeats can exist in a state that is partially structured, requiring relatively low force to unfold; or in an unstructured state which requires almost no force to unfold (Patra et al., 2017). Antibody accessibility studies also support a model in which CSP is conformationally dynamic undergoing significant conformational changes in the oocyst, mosquito salivary gland, and human liver (Coppi et al., 2011, Herrera et al., 2015). The CSP repeat structure may serve to extend the N-terminal domain away from the plasma membrane of the parasite. It is possible that deletion of the repeat sequence (Ferguson et al., 2014) results in a shorter rod-like structure or alternatively it may prevent the entire protein from folding correctly. Such differences in an important protein on the surface of the sporozoite may be detrimental to development in the mosquito.

Liver stage antigen-1 – a network of covalently crosslinked repeats

Liver stage antigen-1 is localised in the parasitophorous vacuole during the liver stage of the lifecycle and is required for progression of *P. falciparum* through this stage (Fidock et al., 1994, Mikolajczak et al., 2011). The repeats in Liver stage antigen-1 are rich in glutamine and lysine residues and are an *in vitro* substrate for transglutaminase which catalyses the covalent crosslinking of polypeptides. Although transglutaminase-crosslinked peptides can be detected in liver stage parasites by microscopy, it is unclear whether LSA-1 itself is crosslinked *in vivo* (Nicoll et al., 2011), and whether this is important for liver-stage development.

RNA Polymerase – a hub for regulated recruitment of binding partners

In eukaryotes, three different RNA Polymerases (RNA Pol) are required to transcribe the nuclear genome; RNA Pol I transcribes ribosomal RNA (rRNA), RNA Pol II transcribes messenger RNA (mRNA) and RNA Pol III transcribes 5S rRNA and transfer RNA (tRNA) (Vannini and Cramer, 2012). As discussed previously, the C-terminus of the largest subunit (RbpI) of RNA Pol II contains a repetitive sequence known as the C-terminal Domain (CTD) (Corden et al., 1985). This domain consists of a tandemly repeated Tyr-Ser-Pro-Thr-Ser-Pro-Ser/Lys sequence which undergoes reversible phosphorylation during the transcription cycle. Changes in the phosphorylation pattern of repeats within the CTD are thought to orchestrate the recruitment of various nuclear factors that couple processing of the nascent mRNA to transcription (Howe, 2002)(Fig. 3).

The CTD is essential for cell viability in yeast, drosophila and mice (Litingtung et al., 1999, Nonet et al., 1987, Zehring et al., 1988). The overall length or number of the heptad repeats is also important as decreasing the number of repeats in the CTD of yeast from 26 to below 8 causes loss of viability (West and Corden, 1995). Although small truncations of the yeast

CTD are tolerated, these result in reduced transcript levels of certain genes. Truncation of the CTD in yeast to 17, 14, 13 and 11 repeats leads to the production of 58%, 8%, 5% and 2% of normal transcript levels of specific genes, respectively, signifying the importance of the repetitive nature of the CTD (Liao et al., 1991).

Although the consensus sequence and number of heptad repeats in RNA pol II varies from one organism to another, it is strictly conserved along evolutionary lineages; for example all mammalian CTDs possess 52 heptapeptide repeats with little variation in the consensus (Chapman et al., 2008). Strikingly, there is considerable variation in the number of repeats found across the *Plasmodium* genus. Rodent and bird-infecting species, including *P. berghei* and *P. gallinaceum,* respectively, possess 8 repeats, while the CTD of some primate-infecting species has expanded resulting in 13-15 repeats. This expansion appears to have arisen twice during *Plasmodium* evolution; once in the lineage that has given rise to *P. falciparum* and once in the lineage that has given rise to *P. vivax* (Kishore et al., 2009). Primate infecting parasites appear to have preferentially expanded a heptad repeat containing a lysine residue in the 7th position (Kishore et al., 2009). Furthermore, this plasticity extends not only to different *Plasmodium* species but also to different isolates within a given species, making *Plasmodium* one of the few organisms that show intra-species variation in the number of heptad repeats in its RNA pol II CTD (Kishore et al., 2009). Although the significance of the repeat expansions is unclear it has been suggested that repeat expansion may confer additional functions to the CTD (Ukaegbu and Deitsch, 2015).

In higher eukaryotes, it is known that the CTD recruits proteins with diverse roles in transcription and gene expression. These include factors such as histone modifying enzymes (Yoh et al., 2008). The CTD of *P. falciparum* RNA pol II interacts with PfSET2 (also referred to as PfSETvs), a histone methyl transferase which tri-methylates histone H3 at lysine 36 (H3K36me3) (Cui et al., 2008, Jiang et al., 2013, Ukaegbu et al., 2014). This histone modification is found in telomeric and subtelomeric chromatin and associated with genomic regions encoding the PfEMP1 *var* gene family (Jiang et al., 2013, Ukaegbu et al., 2014). Although there are more than 60 *var* genes in the *P. falciparum* genome, mRNA from only a single *var* gene is produced at any one time and consequently a single PfEMP1 protein is displayed on the surface of an infected erythrocyte. By switching *var* gene expression the parasite is able to evade an immune response directed against a particular PfEMP1 protein. Although mRNA from only a single *var* gene is transcribed at any one time the RNA pol II also transcribes non-coding RNAs from *var* genes. It is likely that RNA pol II recruits PfSET2 to *var* genes during the latter process. Although the precise mechanism by which PfSET2 and H3K36 tri-methylation regulates *var* gene silencing is unclear, deletion of PfSET2 or expression of a dominant-negative version of PfSET2 which interacts with the RNA pol II CTD, result in dis-regulation of *var* gene expression, strongly indicating a role for the CTD and its interacting partner PfSET2 in this process (Jiang et al., 2013, Ukaegbu et al., 2014).

Homopolymeric Asparagine-rich repeats and protein aggregation

One of the best studied examples of repetitive protein sequences modulating protein properties comes from the study of poly-glutamine (Gln) tracts. Expansion of these homopolymeric tracts leads to protein aggregation and is associated with ‘poly-glutamine diseases’ such as Huntington’s disease in humans (Gatchel and Zoghbi, 2005). However, it has also been proposed that such poly-Gln repetitive sequences modulate protein function and that expansion and contraction of repeats may fine-tune protein function and phenotype over evolutionary time (Gemayel et al., 2015).

Single amino acid repeats of the amino acid asparagine (Asn) are prevalent in the *P. falciparum* genome. Like poly-Glutamine sequences, poly-Asn sequences are also prone to aggregation particularly at elevated temperatures (Halfmann et al., 2011) which the parasite would experience during fever. Given that protein aggregation is often toxic to cells it is remarkable that the parasite proteome maintains so many potentially toxic Asn-rich repeats; more than a quarter of its proteins contain such sequences. Expression of a GFP-tagged version of the Asn-rich protein from *P. falciparum,* PF3D7_0923500, in human cells leads to accumulation of fluorescent protein aggregates as does expression of the N-terminal prion forming domain of *S. cereviasiae* asparagine/glutamine-rich Sup35 (Muralidharan et al., 2012). Co-expression of the human Hsp110 or *Plasmodium* homologue PfHsp110c inhibits aggregate formation; PfHsp110c does so approximately 10-15 fold more efficiently than the human chaperone (Muralidharan et al., 2012). This suggests that *Plasmodium* chaperones may be particularly efficient at suppressing the aggregation of Asn-rich repeat sequences. Although PfHsp110c is essential for survival, a reduction of expression leads to increased aggregation of Asn-rich proteins and parasites cannot recover from short periods of heat shock (Muralidharan et al., 2012). These data are consistent with a model in which PfHsp110c plays a key role in maintaining the solubility of Asn-rich proteins under heat shock conditions (Muralidharan et al., 2012).

The function of Asn-rich repeats in the parasite is unclear and in many cases these sequences have likely evolved without cost or benefit. Notably, deletion of the Asn-rich repeat sequence in the proteasome protein Rpn6 does not compromise its function (Muralidharan et al., 2011). However, several transcription factors in *S. cerevisiae* contain poly-glutamine repeats and variation of the repeat length can fine-tune transcription of target genes and resulting phenotypes (Gemayel et al., 2015). This is at least partially due to changes in protein solubility and interaction partners. Notably, the majority of AP-2 transcription factors encoded in the parasite genome have very high asparagine content of between 20% and 30%. The *P. falciparum* protein AP2-g, which triggers the transition to gametocytogenesis (Kafsack et al., 2014, Sinha et al., 2014), comprises 2432 residues, 29.9% of which are asparagine. Asparagines are found both in homopolymeric repeats, the longest of which is 33 residues in length, as well as in short tandemly repeated sequences. It is unclear if variations in the AP2-g Asn-rich repeats seen in different strains (Campino et al., 2016) fine tune its solubility/activity and hence the propensity of parasites to form gametocytes. Changes in Poly-Asn sequences in other AP2 transcription factors might also modulate their activity leading to small changes in progression through the lifecycle; even small delays in the asexual cycle of the parasite could lead to a selective advantage. Additionally, given the propensity of asparagine-rich repeats to aggregate, particularly under heat shock, it is possible that solubility/activity of transcription factors might change during fever.

Asparagine repeats and protein translation

The rate of translation of a protein is determined by the concentration of available aminoacylated tRNAs. Given the unusually high occurrence of asparagine residues in the *P. falciparum* genome, particularly in long low complexity repetitive sequences, the parasite might require large amounts of Asn:tRNA to efficiently translate Asn-rich proteins. In many organisms there is redundancy in the genes encoding tRNAs for each amino acid. For example, in *C. elegans* there are 46 genes encoding tRNAs that recognise the four glycine codons (Duret, 2000). Given the abundance of Asn in encoded proteins, it might be expected that the *P. falciparum* genome would encode multiple tRNAs that recognize Asn codons thus allowing the parasite to make high levels of Asn:tRNA. However, only a single gene encodes Asn:tRNA (Gardner et al., 2002) and levels of Asn:tRNA are not proportionally more

abundant. This has led to a model in which translation is slowed down when the ribosome encounters long low complexity Asn-repeat sequences due to the relatively lower concentrations of Asn-tRNA (Filisetti et al., 2013, Frugier et al., 2010). This could slow the rate of polypeptide emergence from the ribosome tunnel and allow N-terminal domains that have already been synthesised more time to fold before emergence of subsequent domains. This may increase the overall efficiency of protein folding.

MODIFICATION OF THE HOST ERYTHROCYTE BY REPETITIVE EXPORTED PROTEINS

The human erythrocyte is a relatively simple terminally differentiated cell that lacks internal organelles. During the blood stage of infection the parasite profoundly remodels its host cell, establishing new permeability pathways for acquiring nutrients, modifying the erythrocyte cytoskeleton to increase rigidity, and generating knob-like structures on the surface of the host cell which allow infected cells to cytoadhere to endothelial cells in the microvasculature (Boddey and Cowman, 2013, de Koning-Ward et al., 2016, Przyborski et al., 2016, Spielmann and Gilberger, 2015, Spillman et al., 2015). The parasite brings about these changes by exporting over 400 proteins into the host cell; these proteins can be localised to the erythrocyte cytoplasm, plasma membrane, the underlying spectrin cytoskeleton, or to Maurer's clefts which are membranous structures assembled by the parasite (Fig. 4). Many exported proteins are characterized by the presence of an N-terminal signal sequence followed by a short motif referred to as a Protein Export Element (PEXEL) or Host targeting motif (HT) (Boddey et al., 2013, Hiller et al., 2004, Marti et al., 2004, Schulze et al., 2015). Another class of exported proteins lack a PEXEL sequence and are referred to as PEXEL negative exported proteins (PNEPS) (Heiber et al., 2013). Notably, many exported proteins contain repetitive sequences that play key roles in diverse processes including cytoadhesion, formation of Maurer's clefts, protein localisation, and detoxification of heme (Fig. 4).

Lysine-rich repeats in exported proteins – Changes in repeat length can influence protein localization

There are several examples of repetitive proteins, such as the yeast proteins Flo1p and Flo11p (discussed in previous section), where the function of the protein can be proportionally modulated by repeat length (Fidalgo et al., 2006, Verstrepen et al., 2005). It is conceptually clear in such cases how relatively small changes in the number of tandemly repeated sequences could be important in facilitating adaption of an organism to its environment. In *Plasmodium*, many proteins including exported proteins, contain low complexity lysine-rich repetitive sequences. In some cases these sequences are adjacent to a folded domain but they can also be found in proteins that are predicted to be entirely intrinsically disordered. In several proteins which are exported into the host cell by a PEXEL motif, lysine-rich sequences can target the protein to the periphery of the infected erythrocyte (Davies et al., 2016). Comprising relatively simple tandemly repeated motifs of as few as three residues (e.g. Lys-Lys-Glu), these sequences are likely to interact with either lipids or proteins in the erythrocyte plasma membrane or the underlying erythrocyte spectrin cytoskeleton.

The exported *P. falciparum* protein glutamic acid-rich protein (GARP) contains multiple lysine-rich sequences that function in this manner. Significantly, the targeting propensity of a repetitive lysine-rich sequence from GARP is proportional to its length - increasing the number of tandemly repeated sequences proportionally shifts the protein from being diffusely

localised in the erythrocyte cytoplasm to being highly concentrated adjacent to the plasma membrane of the infected cell (Fig. 5) (Davies et al., 2016). At least nine exported parasite proteins contain lysine-rich sequences that independently target to the periphery of the erythrocyte. The lengths of repeating lysine-rich sequences from many of these exported proteins vary between parasite strains (Davies et al., 2016). Additionally, in the protein MESA, a block of sequence made up of tandemly repeated lysine-rich motifs has itself been duplicated with two to four copies of the sequence in different parasite strains (Davies et al., 2016). Although the phenotypic consequences of repeat expansion and duplication have not been determined experimentally, several of these proteins are known to modulate rigidity and adhesion of infected cells. Given this, expansion and contraction of lysine-rich sequences may proportionally change the concentration of proteins localised to the periphery of the host cell, fine-tuning processes such as cell rigidity or adhesion, and facilitating adaption of the parasite to selective pressure over evolutionary time.

A comparison of the protein sequence of GARP from three different species*, P. falciparum, P. reichenowi,* and *P. gaboni,* indicates that repeat expansion can also lead to *de novo* formation of a targeting domain (Davies et al., 2016). There are three stretches of lysine-rich repetitive sequence within *P. falciparum* GARP. The first sequence in the *P. falciparum* protein is able to target a GFP reporter to the periphery of the host cell. However, the equivalent lysine-rich sequence in the *P. reichenowi* protein is shorter and does not efficiently target to the periphery of the host cell. The second lysine-rich sequence is conserved and is likely to be functional in all three species. The third lysine-rich sequence, that comprises two different repeated motifs in *P. falciparum* GARP, can target a protein to the erythrocyte periphery but the equivalent sequence in *P. gaboni* comprises only one or two repeated lysine-rich motifs and cannot target GFP to the periphery of the infected erythrocyte. These observations show not only that repeat expansion of lysine-rich sequences can modulate protein localization but also that functional targeting sequences can be generated *de novo* by repeat expansion over evolutionary time.

The localisation of Hyp12 (PF3D7_1301400), which modulates the rigidity of the infected cell (Maier et al., 2008), appears to be defined by two conflicting sequences; a repetitive lysine-rich sequence and a repetitive acidic sequence (Davies et al., 2016). Hyp12 is unusual in that the lysine-rich repeat sequence alone is able to target GFP to the periphery of the infected cell but the GFP-tagged full-length protein is localised in the erythrocyte cytoplasm (Davies et al., 2016, Petersen et al., 2015). Removal of the acidic sequence results in the protein being localised to the red cell periphery suggesting that the acidic sequence inhibits the targeting function of the lysine-rich repeat sequence. It is possible that a specific stimulus might unmask the sequence during an infection. Over evolutionary time, it is likely that either expansion of the lysine-rich region or contraction of the acidic sequence could shift the protein from the erythrocyte cytoplasm to the periphery of the erythrocyte.

Knob-associated histidine-rich protein – repeat protein interactions and parasite virulence

Knob-associated histidine-rich protein (KAHRP) is one of the best-characterised repeat-containing proteins. It is a key component of structured spiral-like ‘knob’ protrusions which the parasite assembles at the surface of the infected erythrocyte (Crabb et al., 1997, Watermeyer et al., 2016). Knobs are involved in displaying PfEMP1 proteins on the surface of the infected cell where they mediate adhesion to specific ligands allowing parasites to become sequestered in tissues thus avoiding passage through the spleen (Crabb et al., 1997). KAHRP interacts with the host cytoskeletal components spectrin (Kilejian et al., 1991, Pei et

al., 2005) and ankyrin (Weng et al., 2014) and deletion of the gene causes a loss of knob structures, decreases adhesion under flow conditions (Crabb et al., 1997), and reduces erythrocyte rigidity (Glenister et al., 2002). KAHRP contains an N-terminal repetitive poly-histidine stretch that will target GFP to the periphery of the infected erythrocyte (Rug et al., 2006). It also has two unique lysine-rich repetitive regions, known as the 5' and 3' repeats; only the 5' repeat sequence is sufficient to independently target GFP to the erythrocyte periphery (Davies et al., 2016). Progressive truncation of the C-terminus of KAHRP indicates that the 5' repeats are important for the role of KAHRP in cytoadhesion (Rug et al., 2006). Several studies indicate that these repeats interact with spectrin (Cutts et al., 2017, Kilejian et al., 1991, Pei et al., 2005). Most recently, it has been demonstrated *in vitro* that the repeats interact with β-spectrin domains 10-14 and more weakly with α-spectrin domains 12-16. Significantly, progressive removal of repeats from the KAHRP repeat sequence leads to a corresponding reduction of binding affinity for β-spectrin domains 10-14 i.e. the binding affinity is proportional to the number of KAHRP repeats (Fig. 5) (Cutts et al., 2017). This again highlights the potential importance of repeat expansion and contraction in modulating protein function. Modelling and molecular dynamics simulations support a model whereby the KAHRP repeats adopt an extended conformation of the surface of spectrin, interacting with charge complementary surfaces on multiple β-spectrin domains (Cutts et al., 2017).

Although knob structures are unique to *P. falciparum,* KAHRP-like proteins are found in all primate-infecting *Plasmodium* species (Davies et al., 2016, Sargeant et al., 2006). These proteins share an N-terminal EKAL domain (EMP3 KAHRP-like domain) followed by lysine-rich repeating sequences but the function of the proteins in other species is not known. The repeated lysine-rich motifs in the *P. falciparum* and *P. knowlesi* proteins both contain consecutive lysine residues but otherwise differ significantly. However, when expressed in *P. falciparum*, both can target a protein to the periphery of the infected cell (Davies et al., 2016). It is likely that both these repetitive lysine-rich sequences interact with a negatively charged surface on spectrin. Presumably the binding interface can tolerate somewhat diverse conformations of positively charged interacting peptides. Notably, there does not appear to be a strong consensus sequence among all of the different lysine-rich targeting sequences identified thus far. Identification of the binding partners, which could be proteins or lipids, for each protein and further structural studies will be required to understand the functions of these diverse repetitive sequences.

Although KAHRP contains a conserved EKAL domain of unknown function at its N-terminus (Davies et al., 2016, Sargeant et al., 2006), the remainder of the protein is predicted to be intrinsically disordered and much of the sequence is repetitive. Likewise, PfEMP3 contains a conserved N-terminal EKAL domain and several different repetitive sequences at its C-terminus but the repetitive sequences are very different from those of KAHRP. It is likely that the PfEMP3 gene arose as a duplication of the KAHRP gene or vice versa yet gene knockout and functional studies show that the two proteins are not functionally equivalent (Crabb et al., 1997, Glenister et al., 2002, Waterkeyn et al., 2000). Likely, after gene duplication the dynamic repetitive sequences have diverged dramatically and consequently the function of the two proteins has likewise also diverged. The dynamic nature of repetitive sequences together with the telomeric location of many of the genes encoding exported proteins suggests that gene duplication and rapid diversification have occurred frequently during the evolution of the parasite; repeat expansion likely plays an important role in the latter process.

PHIST/PRESAN domain proteins

A large group of functionally diverse exported parasite proteins belong to the PHIST (Plasmodium helical interspersed subtelomeric) or PRESAN (Plasmodium RESA N-terminal) domain family (Oakley et al., 2007, Sargeant et al., 2006, Warncke et al., 2016). The proteins are characterized by the presence of a conserved alpha helical domain (Mayer et al., 2012) and are exported into the infected erythrocyte during the asexual erythrocyte stages and during gametocytogenesis (Oakley et al., 2007, Sargeant et al., 2006, Silvestrini et al., 2010, Tiburcio et al., 2015). One of the best characterised proteins in this family is LyMP (lysine-rich membrane-associated PHISTb), which modulates erythrocyte cytoadhesion by linking PfEMP1 to the erythrocyte cytoskeleton (Oberli et al., 2014, Oberli et al., 2016, Proellocks et al., 2014, Tarr et al., 2014). LyMP contains an N-terminal PHIST/PRESAN domain, which interacts with the cytoplasmic tail of PfEMP1 (Oberli et al., 2014), and a C-terminal sequence that interacts with the erythrocyte cytoskeletal protein Band 3 (Oberli et al., 2016). The extreme C-terminus of LyMP contains a lysine-rich repetitive sequence which when GFP-tagged and exported into the erythrocyte is sufficient to localise to the erythrocyte periphery (Davies et al., 2016). Although the fragment of LyMP that has been shown to interact with Band 3 *in vitro* contains the lysine-rich sequence, it is unclear whether this sequence interacts with Band 3 or another component of the erythrocyte cytoskeleton.

LyMP is an example of a gene that encodes a folded domain (PHIST/PRESAN domain) but has expanded a repetitive lysine-rich sequence which likely allows the encoded protein to interact with an additional protein. Similarly, GEXP12, PF3D7_1201000, PF3D7_1476200, and PF3D7_0402000 contain a PHIST/PRESAN domain but also contain lysine-rich sequences that alone localise to the periphery of the infected erythrocyte (Davies et al., 2016). The PHIST/PRESAN domain of PF3D7_0402000 interacts with Band 4.1 and ankyrin that are components of the erythrocyte cytoskeleton (Parish et al., 2013, Shakya et al., 2017, Vignali et al., 2008), but the protein also contains a C-terminal lysine-rich repetitive targeting sequence (Davies et al., 2016). Other PHIST/PRESAN proteins, such as PFD80 (Tarr et al., 2014) and RESA (Ring expressed surface antigen) (Brown et al., 1985) localize to the erythrocyte cytoskeleton, and have expanded repetitive sequences that are not lysine-rich and do not independently confer a cytoskeletal localization. RESA modulates the rigidity of the infected cell and deletion of the gene does not affect cytoadhesion (Maier et al., 2008). Although the role of PFD80 is unclear it is likely essential as attempts to delete the gene have been unsuccessful (Maier et al., 2008). The localization and likely functions of PHIST proteins are diverse and in most cases the functions of their repetitive sequences are unclear. However, as with KAHRP and PfEMP3, it is possible that the functions of the protein family members have been diversified by duplication of genes encoding a protein with a conserved domain followed by expansion of different repetitive sequences which may contribute to modulating or changing the function of the protein.

Repetitive Maurer's cleft proteins

There are numerous exported proteins with repetitive sequences that are not lysine-rich. In some cases the function of these sequences is well understood. Ring exported protein-1 (REX1) is a PEXEL negative exported protein which has a repetitive sequence that is essential for protein function (McHugh et al., 2015). The protein is peripherally associated with Maurer's clefts (Hawthorne et al., 2004) which are flattened membrane vesicles in the infected erythrocyte and are essential for the trafficking of PfEMP1 to the surface of the infected erythrocyte. Deletion of genes encoding several Maurer's cleft proteins disrupts PfEMP1 trafficking (Cooke et al., 2006, Maier et al., 2007, Rug et al., 2014, Spycher et al., 2008).

Deletion of the repetitive region of REX1 (residues 371 - 579) reduces the number of Maurer's clefts detected in the erythrocyte. On average 13 punctae corresponding to Maurer's clefts are seen in erythrocytes infected with wildtype parasites but only 2 large punctae are seen in cells infected with parasites expressing REX1 that is missing the repetitive sequence (McHugh et al., 2015). Electron microscopy reveals that these Maurer's clefts have an unusual stacking phenotype with an average of six clefts stacked on top of each other. (McHugh et al., 2015). Although PfEMP1 is trafficked to the abnormal stacked Maurer's clefts, transfer to the erythrocyte plasma membrane is less efficient when REX1 is either knocked down or the repetitive region is deleted (McHugh et al., 2015). Removal of the repetitive region also reduces the adhesion of infected erythrocytes to CD36 cells under flow conditions (McHugh et al., 2015). This indicates that the repetitive region of REX1 is functionally important for both Maurer's cleft architecture and efficient trafficking of PfEMP1. It is unlikely that the repeat region is simply forming a linker sequence between the N- and C-termini of the protein, as the C-terminal residues are not required for protein function and can be deleted. This suggests that the repeat sequences play an active role in protein function, possibly forming a key structural element or recruiting other proteins to the Maurer's cleft membrane. Notably, the sequence and length of the repetitive domain varies between different parasites; this could contribute to the formation of stacked cleft structures in some strains (Wickert and Krohne, 2007) and may modulate the propensity of parasites to cytoadhere (McHugh et al., 2015).

Skeleton binding protein-1 (SBP1) is an exported PNEP which contains a single transmembrane segment which is integrated within the Maurer's cleft membrane. Deletion of the gene results in reduced trafficking of PfEMP1 to the erythrocyte surface and reduced cytoadhesion (Chan et al., 2016, Cooke et al., 2006, Maier et al., 2007). SBP1 contains repetitive sequences located N- and C-terminally on either side of the transmembrane segment. The N-terminal repeats can self-associate (Saridaki et al., 2009) while the glutamine-rich repeats at its C-terminus associate with the cytoskeletal components spectrin and 4.1 *in vitro* (Kats et al., 2015). Although these interactions likely contribute to the tethering of Maurer's clefts to the erythrocyte cytoskeleton, the C-terminus of SBP1 can be replaced by an unrelated protein sequence without altering the location of Maurer's clefts or cytoadhesion of infected cells. However, as it is known that there are multiple mechanisms for tethering of Maurer's clefts there may be redundancy in this function of SBP1 (Cyrklaff et al., 2011, Hanssen et al., 2010, Pachlatko et al., 2010, Rug et al., 2014).

Pf332 is a large exported protein that localises to Maurer's clefts and is mainly composed of several different glutamate-rich repeats. Pf332 deletion and truncation of its repetitive C-terminus have been shown to have varying effects on the rigidity and cytoadhesion of infected cells; differences between studies may be a result of using different parasite strains and the particular deletion strategies (Glenister et al., 2009, Hodder et al., 2009).

Exported Repetitive histidine-rich proteins

Histidine-rich Protein 2 (HRP2) is a *P. falciparum* protein comprised almost entirely of repetitive Histidine-rich sequences. HRP2 is exported into the erythrocyte cytoplasm during the blood stage of infection and released into the blood during parasite egress (Howard et al., 1986) where it can accumulate to high levels (Ramutton et al., 2012). Several functions for HRP2 have been proposed including the detoxification of heme. The parasite internalizes haemoglobin from the cytoplasm of the host erythrocyte and digests it within the food vacuole; a fraction of HRP2 is also localised in the food vacuole (Papalexis et al., 2001,

Sullivan et al., 1996). Within the food vacuole the toxic heme moiety, a by-product of haemoglobin degradation, is converted into inert hemozoin. *In vitro*, HRP2 is able to catalyse the conversion of heme to hemozoin as is a similar histidine-rich repetitive protein HRP3 (Sullivan et al., 1996). HRP2 is also able to promote the degradation of heme (Papalexis et al., 2001). However, *P. falciparum* parasites lacking genes encoding both HRP2 and HRP3 are able to produce hemozoin suggesting that there are multiple mechanisms by which hemozoin is generated *in vivo* or that HRP2 plays a relatively minor role in this process. Indeed, other protein-based (Jani et al., 2008, Sullivan et al., 1996) and non-protein based mechanisms have been proposed (reviewed in (Egan, 2008)).

HRP2 has also been implicated in regulation of thrombin inhibition (Ndonwi et al., 2011). *In vitro,* anti-thrombin inhibits the activity of thrombin and factor 10a and the inhibition is potentiated by glycosaminoglycans such as heparin. The potentiation can be antagonized by the interaction of HRP2 with glycosaminoglycans (Ndonwi et al., 2011). Thus HRP2 may behave as a pro-coagulant during infection.

More recent studies indicate that HRP2 may play a role in cerebral malaria by disrupting the integrity of brain endothelial cell barriers. Using a system that comprises an *in vitro* cultured monolayer of human endothelial cells which mimics the blood brain barrier, it has been shown that HRP2 can disrupt the integrity of an endothelial barrier (Pal et al., 2016). Addition of parasites that produce HRP2 (but not parasites that do not make HRP2) or recombinant HRP2 protein to the monolayer of endothelial cells results in redistribution of tight junction and adherens junction proteins, and disruption of the barrier formed by the cell monolayer. Blocking of key signalling factors in the host cell prevents the disruption caused by HRP2 supporting a model of inflammasome mediated activation of the innate immune system in this process. Exposure to HRP2 also increases the number of VCAM-1 and ICAM-1 receptors on the surface of endothelical cells. Both receptors can be used by infected erythrocytes to adhere to endothelial cells suggesting that activation of an inflammatory response may enhance cytoadherence and sequestration of parasites (Pal et al., 2016). Although mouse malaria parasites lack an HRP2 homolog, injection of HRP2 protein reduces survival of infected mice (Pal et al., 2017). It remains to be determined whether the incidence of cerebral malaria in humans is lower in infections caused by HRP2-negative parasites.

Maurer's cleft associated histidine-rich protein 1 (MAHRP-1) is an exported protein localized to Maurer's clefts. The protein contains a single transmembrane segment and a repetitive histidine-rich C-terminus, and is involved in protein trafficking within the infected erythrocyte (Spycher et al., 2008). Like HRP2, the repetitive histidine-rich sequence of MAHRP is implicated in heme detoxification (Papalexis et al., 2001). The repetitive sequence is able to bind ferric heme, which is toxic to cells, and accelerates its degradation mediated by hydrogen peroxide. This suggests that the histidine-rich sequence of MAHRP-1 may play a role in scavenging ferric heme and detoxification of reactive oxygen species within the infected erythrocyte (Spycher et al., 2003). Notably, the binding stoichiometry between ferric heme and MAHRP-1 is proportional to the number of repeats within the protein. This suggests that the potential protective function of MAHRP-1 could be influenced by repeat expansion (Spycher et al., 2003).

Histidine-rich repetitive proteins and diagnosis of malaria

The World Health Organisation now recommends diagnostic testing of people with suspected malaria before treatment (WHO, 2016). Rapid diagnostic tests (RDTs) represent an effective tool in this context. Most RDTs utilize antibodies to detect the HRP2 protein in blood. The mature protein, after cleavage of the PEXEL/HT motif (Chang et al., 2008), comprises

histidine-rich repeated sequences which form the epitopes recognized in RDTs (Wellems and Howard, 1986). More than 20 sequence variants of these repeat sequences have been identified (Baker et al., 2010, Deme et al., 2014). There is significant polymorphism in overall length, sequence composition of the repeated motifs, and the number of repeated motifs, between different *P. falciparum* strains (Baker et al., 2010, Baker et al., 2005, Deme et al., 2014, Ramutton et al., 2012) but there is currently no indication that sequence polymorphism can cause false negative RDT tests (Baker et al., 2010, Ramutton et al., 2012). However, increasing RDT use may place a selective pressure on the HRP2 gene and a contraction or mutation of repeats recognised by antibodies in RDTs may result in reduced test sensitivity or failure. Parasites lacking the HRP2 gene, initially identified in Peru (Gamboa et al., 2010) and subsequently other regions, do yield negative HRP2 based RDT tests (Bharti et al., 2016, Amoah et al., 2016, Li et al., 2015, Murillo Solano et al., 2015, Akinyi Okoth et al., 2015, Abdallah et al., 2015, Kumar et al., 2013, Wurtz et al., 2013, Maltha et al., 2012, Houze et al., 2011, Koita et al., 2012, Dorado et al., 2016, Baldeviano et al., 2015, Cheng et al., 2014).

CONCLUSION

In summary, repetitive sequences are widespread in *Plasmodium* genomes (Mendes et al., 2013) and are particularly abundant in *P. falciparum* genes. These dynamic sequences can arise, expand, or contract rapidly. Due to the assumption that they represent non-functional 'junk DNA' and are challenging to accurately assemble from short-read sequence data, they are mostly excluded from genomic analyses. Indeed, it is reasonable to assume in many cases that repeats simply expand and contract rapidly due to frequent errors in DNA replication, and evolve neutrally. Although it is difficult to provide a universal explanation for the presence of these sequences in the parasite genome, it appears that relatively simple homopolymeric repeats tend to evolve neutrally but more complex sequences appear to be under selection suggesting that they are functional (Battistuzzi et al., 2016, Zilversmit et al., 2010).

The fraction of *Plasmodium* repetitive protein sequences that are functionally important remains unclear and most remain uncharacterized. It is evident that some sequences within open reading frames encode repetitive polypeptide sequences that can modulate protein activity or are integral to protein function forming key structural elements, binding sites for other proteins and ligands, and localisation signals. Repeat expansion and contraction can likely modulate and diversify protein function resulting in gain or loss of function phenotypes that allow the parasite to adapt under selective pressure. Despite the diverse functions of repetitive sequences it is conceptually clear in many cases how changes in repeat length might influence protein function and parasite phenotype. The development of new genetic tools will now allow the importance of these sequences to be tested. In the simplest cases, relatively small changes in length of diverse repeat types may alter protein folding efficiency, stability, or aggregation leading to selective advantage for a parasite. In cases such as the CTD of RNA Polymerase II, where repeat sequences have very specific functions in the regulated recruitment of protein binding partners, changes in repeat number may fine tune protein function in a manner proportional to the number of repeats. The abundance of repetitive sequences that are recognized by the immune system and the occurrence of functional repetitive sequences in proteins such as KAHRP and REX1 that are central players in the pathology of severe malaria, highlight why an understanding of these sequences is important to our understanding of severe disease, its treatment, and prevention.

FUNDING:

This work was supported by the Wellcome Trust [099764 Z/12/Z] the Biotechnological and Biological Sciences research council [BB/M009513/1].

Footnote:

Corresponding author: Andrew Osborne

[1]- Institute of Structural and Molecular Biology, Department of Biological Sciences, Birkbeck and University College London, London, UK
E-mail a.osborne@ucl.ac.uk; Telephone (+44) 0207 679 3155; Fax (+44) 0207 679 7046.
[*] - authors contributed equally
[2] - current address - The Francis Crick Institute, London, United Kingdom, UK
[3] - current address - London School of Hygiene & Tropical Medicine, London, UK

ABDALLAH, J. F., OKOTH, S. A., FONTECHA, G. A., TORRES, R. E., BANEGAS, E. I., MATUTE, M. L., BUCHELI, S. T., GOLDMAN, I. F., DE OLIVEIRA, A. M., BARNWELL, J. W. & UDHAYAKUMAR, V. 2015. Prevalence of pfhrp2 and pfhrp3 gene deletions in Puerto Lempira, Honduras. *Malar J,* 14**,** 19.

AKINYI OKOTH, S., ABDALLAH, J. F., CERON, N., ADHIN, M. R., CHANDRABOSE, J., KRISHNALALL, K., HUBER, C. S., GOLDMAN, I. F., MACEDO DE OLIVEIRA, A., BARNWELL, J. W. & UDHAYAKUMAR, V. 2015. Variation in *Plasmodium falciparum* Histidine-Rich Protein 2 (Pfhrp2) and *Plasmodium falciparum* Histidine-Rich Protein 3 (Pfhrp3) Gene Deletions in Guyana and Suriname. *PLoS One,* 10**,** e0126805.

ALY, A. S., VAUGHAN, A. M. & KAPPE, S. H. 2009. Malaria parasite development in the mosquito and infection of the mammalian host. *Annu Rev Microbiol,* 63**,** 195-221.

AMOAH, L. E., ABANKWA, J. & OPPONG, A. 2016. *Plasmodium falciparum* histidine rich protein-2 diversity and the implications for PfHRP 2: based malaria rapid diagnostic tests in Ghana. *Malar J,* 15**,** 101.

ANDRADE, M. A., PEREZ-IRATXETA, C. & PONTING, C. P. 2001. Protein repeats: structures, functions, and evolution. *J Struct Biol,* 134**,** 117-31.

BAKER, J., HO, M. F., PELECANOS, A., GATTON, M., CHEN, N., ABDULLAH, S., ALBERTINI, A., ARIEY, F., BARNWELL, J., BELL, D., CUNNINGHAM, J., DJALLE, D., ECHEVERRY, D. F., GAMBOA, D., HII, J., KYAW, M. P., LUCHAVEZ, J., MEMBI, C., MENARD, D., MURILLO, C., NHEM, S., OGUTU, B., ONYOR, P., OYIBO, W., WANG, S. Q., MCCARTHY, J. & CHENG, Q. 2010. Global sequence variation in the histidine-rich proteins 2 and 3 of *Plasmodium falciparum*: implications for the performance of malaria rapid diagnostic tests. *Malar J,* 9**,** 129.

BAKER, J., MCCARTHY, J., GATTON, M., KYLE, D. E., BELIZARIO, V., LUCHAVEZ, J., BELL, D. & CHENG, Q. 2005. Genetic diversity of *Plasmodium falciparum* histidine-rich protein 2 (PfHRP2) and its effect on the performance of PfHRP2-based rapid diagnostic tests. *J Infect Dis,* 192**,** 870-7.

BALDEVIANO, G. C., OKOTH, S. A., ARROSPIDE, N., GONZALEZ, R. V., SANCHEZ, J. F., MACEDO, S., CONDE, S., TAPIA, L. L., SALAS, C., GAMBOA, D., HERRERA, Y., EDGEL, K. A., UDHAYAKUMAR, V. & LESCANO, A. G. 2015. Molecular Epidemiology of *Plasmodium falciparum* Malaria Outbreak, Tumbes, Peru, 2010-2012. *Emerg Infect Dis,* 21**,** 797-803.

BATTISTUZZI, F. U., SCHNEIDER, K. A., SPENCER, M. K., FISHER, D., CHAUDHRY, S. & ESCALANTE, A. A. 2016. Profiles of low complexity regions in Apicomplexa. *BMC Evol Biol,* 16**,** 47.

BEESON, J. G., DREW, D. R., BOYLE, M. J., FENG, G., FOWKES, F. J. & RICHARDS, J. S. 2016. Merozoite surface proteins in red blood cell invasion, immunity and vaccines against malaria. *FEMS Microbiol Rev,* 40**,** 343-72.

BENNINK, S., KIESOW, M. J. & PRADEL, G. 2016. The development of malaria parasites in the mosquito midgut. *Cell Microbiol,* 18**,** 905-18.

BHARTI, P. K., CHANDEL, H. S., AHMAD, A., KRISHNA, S., UDHAYAKUMAR, V. & SINGH, N. 2016. Prevalence of pfhrp2 and/or pfhrp3 Gene Deletion in *Plasmodium falciparum* Population in Eight Highly Endemic States in India. *PLoS One,* 11**,** e0157949.

BINGHAM, R. J., RUDINO-PINERA, E., MEENAN, N. A., SCHWARZ-LINEK, U., TURKENBURG, J. P., HOOK, M., GARMAN, E. F. & POTTS, J. R. 2008. Crystal structures of fibronectin-binding sites from Staphylococcus aureus FnBPA in complex with fibronectin domains. *Proc Natl Acad Sci U S A,* 105**,** 12254-8.

BODDEY, J. A., CARVALHO, T. G., HODDER, A. N., SARGEANT, T. J., SLEEBS, B. E., MARAPANA, D., LOPATICKI, S., NEBL, T. & COWMAN, A. F. 2013. Role of plasmepsin V in export of diverse protein families from the *Plasmodium falciparum* exportome. *Traffic,* 14**,** 532-50.

BODDEY, J. A. & COWMAN, A. F. 2013. *Plasmodium* nesting: remaking the erythrocyte from the inside out. *Annu Rev Microbiol,* 67**,** 243-69.

BROWN, G. V., CULVENOR, J. G., CREWTHER, P. E., BIANCO, A. E., COPPEL, R. L., SAINT, R. B., STAHL, H. D., KEMP, D. J. & ANDERS, R. F. 1985. Localization of the ring-infected erythrocyte surface antigen (RESA) of *Plasmodium falciparum* in merozoites and ring-infected erythrocytes. *J Exp Med,* 162**,** 774-9.

CAMPINO, S., BENAVENTE, E. D., ASSEFA, S., THOMPSON, E., DROUGHT, L. G., TAYLOR, C. J., GORVETT, Z., CARRET, C. K., FLUECK, C., IVENS, A. C., KWIATKOWSKI, D. P., ALANO, P., BAKER, D. A. & CLARK, T. G. 2016. Genomic variation in two gametocyte non-producing *Plasmodium falciparum* clonal lines. *Malar J,* 15**,** 229.

CHAN, J. A., HOWELL, K. B., LANGER, C., MAIER, A. G., HASANG, W., ROGERSON, S. J., PETTER, M., CHESSON, J., STANISIC, D. I., DUFFY, M. F., COOKE, B. M., SIBA, P. M., MUELLER, I., BULL, P. C., MARSH, K., FOWKES, F. J. & BEESON, J. G. 2016. A single point in protein trafficking by *Plasmodium falciparum* determines the expression of major antigens on the surface of infected erythrocytes targeted by human antibodies. *Cell Mol Life Sci,* 73**,** 4141-58.

CHANG, H. H., FALICK, A. M., CARLTON, P. M., SEDAT, J. W., DERISI, J. L. & MARLETTA, M. A. 2008. N-terminal processing of proteins exported by malaria parasites. *Mol Biochem Parasitol,* 160**,** 107-15.

CHAPMAN, R. D., HEIDEMANN, M., HINTERMAIR, C. & EICK, D. 2008. Molecular evolution of the RNA polymerase II CTD. *Trends Genet,* 24**,** 289-96.

CHENG, Q., GATTON, M. L., BARNWELL, J., CHIODINI, P., MCCARTHY, J., BELL, D. & CUNNINGHAM, J. 2014. *Plasmodium falciparum* parasites lacking histidine-rich protein 2 and 3: a review and recommendations for accurate reporting. *Malar J,* 13**,** 283.

CHERNOVA, T. A., WILKINSON, K. D. & CHERNOFF, Y. O. 2014. Physiological and environmental control of yeast prions. *FEMS Microbiol Rev,* 38**,** 326-44.

COLLINS, C. R., HACKETT, F., ATID, J., TAN, M. S. Y. & BLACKMAN, M. J. 2017. The *Plasmodium falciparum* pseudoprotease SERA5 regulates the kinetics and efficiency of malaria parasite egress from host erythrocytes. *PLoS Pathog,* 13**,** e1006453.

CONWAY, D. J., CAVANAGH, D. R., TANABE, K., ROPER, C., MIKES, Z. S., SAKIHAMA, N., BOJANG, K. A., ODUOLA, A. M., KREMSNER, P. G., ARNOT, D. E., GREENWOOD, B. M. & MCBRIDE, J. S. 2000. A principal target of human immunity to malaria identified by molecular population genetic and immunological analyses. *Nat Med,* 6**,** 689-92.

COOKE, B. M., BUCKINGHAM, D. W., GLENISTER, F. K., FERNANDEZ, K. M., BANNISTER, L. H., MARTI, M., MOHANDAS, N. & COPPEL, R. L. 2006. A Maurer's cleft-associated protein is essential for expression of the major malaria virulence antigen on the surface of infected red blood cells. *J Cell Biol,* 172**,** 899-908.

COPPI, A., NATARAJAN, R., PRADEL, G., BENNETT, B. L., JAMES, E. R., ROGGERO, M. A., CORRADIN, G., PERSSON, C., TEWARI, R. & SINNIS, P. 2011. The malaria

circumsporozoite protein has two functional domains, each with distinct roles as sporozoites journey from mosquito to mammalian host. *J Exp Med,* 208**,** 341-56.
CORDEN, J. L., CADENA, D. L., AHEARN, J. M., JR. & DAHMUS, M. E. 1985. A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II. *Proc Natl Acad Sci U S A,* 82**,** 7934-8.
COWMAN, A. F., HEALER, J., MARAPANA, D. & MARSH, K. 2016. Malaria: Biology and Disease. *Cell,* 167**,** 610-624.
CRABB, B. S., COOKE, B. M., REEDER, J. C., WALLER, R. F., CARUANA, S. R., DAVERN, K. M., WICKHAM, M. E., BROWN, G. V., COPPEL, R. L. & COWMAN, A. F. 1997. Targeted gene disruption shows that knobs enable malaria-infected red cells to cytoadhere under physiological shear stress. *Cell,* 89**,** 287-96.
CUI, L., FAN, Q., CUI, L. & MIAO, J. 2008. Histone lysine methyltransferases and demethylases in *Plasmodium falciparum*. *Int J Parasitol,* 38**,** 1083-97.
CUMBERWORTH, A., LAMOUR, G., BABU, M. M. & GSPONER, J. 2013. Promiscuity as a functional trait: intrinsically disordered regions as central players of interactomes. *Biochem J,* 454**,** 361-9.
CUTTS, E. E., LAASCH, N., REITER, D. M., TRENKER, R., SLATER, L. M., STANSFELD, P. J. & VAKONAKIS, I. 2017. Structural analysis of *P. falciparum* KAHRP and PfEMP1 complexes with host erythrocyte spectrin suggests a model for cytoadherent knob protrusions. *PLoS Pathog,* 13**,** e1006552.
CYRKLAFF, M., SANCHEZ, C. P., KILIAN, N., BISSEYE, C., SIMPORE, J., FRISCHKNECHT, F. & LANZER, M. 2011. Hemoglobins S and C interfere with actin remodeling in *Plasmodium falciparum*-infected erythrocytes. *Science,* 334**,** 1283-6.
DAVIES, H. M., THALASSINOS, K. & OSBORNE, A. R. 2016. Expansion of Lysine-rich Repeats in *Plasmodium* Proteins Generates Novel Localization Sequences That Target the Periphery of the Host Erythrocyte. *J Biol Chem,* 291**,** 26188-26207.
DE KONING-WARD, T. F., DIXON, M. W., TILLEY, L. & GILSON, P. R. 2016. *Plasmodium* species: master renovators of their host cells. *Nat Rev Microbiol,* 14**,** 494-507.
DEME, A. B., PARK, D. J., BEI, A. K., SARR, O., BADIANE, A. S., GUEYE PEL, H., AHOUIDI, A., NDIR, O., MBOUP, S., WIRTH, D. F., NDIAYE, D. & VOLKMAN, S. K. 2014. Analysis of pfhrp2 genetic diversity in Senegal and implications for use of rapid diagnostic tests. *Malar J,* 13**,** 34.
DORADO, E. J., OKOTH, S. A., MONTENEGRO, L. M., DIAZ, G., BARNWELL, J. W., UDHAYAKUMAR, V. & MURILLO SOLANO, C. 2016. Genetic Characterisation of *Plasmodium falciparum* Isolates with Deletion of the pfhrp2 and/or pfhrp3 Genes in Colombia: The Amazon Region, a Challenge for Malaria Diagnosis and Control. *PLoS One,* 11**,** e0163137.
DURET, L. 2000. tRNA gene number and codon usage in the C. elegans genome are co-adapted for optimal translation of highly expressed genes. *Trends Genet,* 16**,** 287-9.
DYSON, H. J., SATTERTHWAIT, A. C., LERNER, R. A. & WRIGHT, P. E. 1990. Conformational preferences of synthetic peptides derived from the immunodominant site of the circumsporozoite protein of *Plasmodium falciparum* by 1H NMR. *Biochemistry,* 29**,** 7828-37.
EDWARDS, A. M., POTTS, J. R., JOSEFSSON, E. & MASSEY, R. C. 2010. *Staphylococcus aureus* host cell invasion and virulence in sepsis is facilitated by the multiple repeats within FnBPA. *PLoS Pathog,* 6**,** e1000964.
EGAN, T. J. 2008. Haemozoin formation. *Mol Biochem Parasitol,* 157**,** 127-36.
ESPINOSA, D. A., YADAVA, A., ANGOV, E., MAURIZIO, P. L., OCKENHOUSE, C. F. & ZAVALA, F. 2013. Development of a chimeric *Plasmodium berghei* strain expressing the repeat region of the *P. vivax* circumsporozoite protein for in vivo evaluation of vaccine efficacy. *Infect Immun,* 81**,** 2882-7.
FENG, Z. P., ZHANG, X., HAN, P., ARORA, N., ANDERS, R. F. & NORTON, R. S. 2006. Abundance of intrinsically unstructured proteins in *P. falciparum* and other apicomplexan parasite proteomes. *Mol Biochem Parasitol,* 150**,** 256-67.

FERGUSON, D. J., BALABAN, A. E., PATZEWITZ, E. M., WALL, R. J., HOPP, C. S., POULIN, B., MOHMMED, A., MALHOTRA, P., COPPI, A., SINNIS, P. & TEWARI, R. 2014. The repeat region of the circumsporozoite protein is critical for sporozoite formation and maturation in *Plasmodium*. *PLoS One,* 9**,** e113923.

FIDALGO, M., BARRALES, R. R., IBEAS, J. I. & JIMENEZ, J. 2006. Adaptive evolution by mutations in the FLO11 gene. *Proc Natl Acad Sci U S A,* 103**,** 11228-33.

FIDOCK, D. A., GRAS-MASSE, H., LEPERS, J. P., BRAHIMI, K., BENMOHAMED, L., MELLOUK, S., GUERIN-MARCHAND, C., LONDONO, A., RAHARIMALALA, L., MEIS, J. F. & ET AL. 1994. *Plasmodium falciparum* liver stage antigen-1 is well conserved and contains potent B and T cell determinants. *J Immunol,* 153**,** 190-204.

FILISETTI, D., THEOBALD-DIETRICH, A., MAHMOUDI, N., RUDINGER-THIRION, J., CANDOLFI, E. & FRUGIER, M. 2013. Aminoacylation of *Plasmodium falciparum* tRNA(Asn) and insights in the synthesis of asparagine repeats. *J Biol Chem,* 288**,** 36361-71.

FISHER, C. R., SUTTON, H. J., KACZMARSKI, J. A., MCNAMARA, H. A., CLIFTON, B., MITCHELL, J., CAI, Y., DUPS, J. N., D'ARCY, N. J., SINGH, M., CHUAH, A., PEAT, T. S., JACKSON, C. J. & COCKBURN, I. A. 2017. T-dependent B cell responses to *Plasmodium* induce antibodies that form a high-avidity multivalent complex with the circumsporozoite protein. *PLoS Pathog,* 13**,** e1006469.

FONDON, J. W., 3RD & GARNER, H. R. 2004. Molecular origins of rapid and continuous morphological evolution. *Proc Natl Acad Sci U S A,* 101**,** 18058-63.

FONDON, J. W., 3RD & GARNER, H. R. 2007. Detection of length-dependent effects of tandem repeat alleles by 3-D geometric decomposition of craniofacial variation. *Dev Genes Evol,* 217**,** 79-85.

FRUGIER, M., BOUR, T., AYACH, M., SANTOS, M. A., RUDINGER-THIRION, J., THEOBALD-DIETRICH, A. & PIZZI, E. 2010. Low Complexity Regions behave as tRNA sponges to help co-translational folding of plasmodial proteins. *FEBS Lett,* 584**,** 448-54.

GALAMO, C. D., JAFARSHAD, A., BLANC, C. & DRUILHE, P. 2009. Anti-MSP1 block 2 antibodies are effective at parasite killing in an allele-specific manner by monocyte-mediated antibody-dependent cellular inhibition. *J Infect Dis,* 199**,** 1151-4.

GAMBOA, D., HO, M. F., BENDEZU, J., TORRES, K., CHIODINI, P. L., BARNWELL, J. W., INCARDONA, S., PERKINS, M., BELL, D., MCCARTHY, J. & CHENG, Q. 2010. A large proportion of *P. falciparum* isolates in the Amazon region of Peru lack pfhrp2 and pfhrp3: implications for malaria rapid diagnostic tests. *PLoS One,* 5**,** e8091.

GARDNER, M. J., HALL, N., FUNG, E., WHITE, O., BERRIMAN, M., HYMAN, R. W., CARLTON, J. M., PAIN, A., NELSON, K. E., BOWMAN, S., PAULSEN, I. T., JAMES, K., EISEN, J. A., RUTHERFORD, K., SALZBERG, S. L., CRAIG, A., KYES, S., CHAN, M. S., NENE, V., SHALLOM, S. J., SUH, B., PETERSON, J., ANGIUOLI, S., PERTEA, M., ALLEN, J., SELENGUT, J., HAFT, D., MATHER, M. W., VAIDYA, A. B., MARTIN, D. M., FAIRLAMB, A. H., FRAUNHOLZ, M. J., ROOS, D. S., RALPH, S. A., MCFADDEN, G. I., CUMMINGS, L. M., SUBRAMANIAN, G. M., MUNGALL, C., VENTER, J. C., CARUCCI, D. J., HOFFMAN, S. L., NEWBOLD, C., DAVIS, R. W., FRASER, C. M. & BARRELL, B. 2002. Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature,* 419**,** 498-511.

GATCHEL, J. R. & ZOGHBI, H. Y. 2005. Diseases of unstable repeat expansion: mechanisms and common principles. *Nat Rev Genet,* 6**,** 743-55.

GEMAYEL, R., CHAVALI, S., POUGACH, K., LEGENDRE, M., ZHU, B., BOEYNAEMS, S., VAN DER ZANDE, E., GEVAERT, K., ROUSSEAU, F., SCHYMKOWITZ, J., BABU, M. M. & VERSTREPEN, K. J. 2015. Variable Glutamine-Rich Repeats Modulate Transcription Factor Activity. *Mol Cell,* 59**,** 615-27.

GEMAYEL, R., VINCES, M. D., LEGENDRE, M. & VERSTREPEN, K. J. 2010. Variable tandem repeats accelerate evolution of coding and regulatory sequences. *Annu Rev Genet,* 44**,** 445-77.

GHASPARIAN, A., MOEHLE, K., LINDEN, A. & ROBINSON, J. A. 2006. Crystal structure of an NPNA-repeat motif from the circumsporozoite protein of the malaria parasite *Plasmodium falciparum*. *Chem Commun (Camb)***,** 174-6.

GLENISTER, F. K., COPPEL, R. L., COWMAN, A. F., MOHANDAS, N. & COOKE, B. M. 2002. Contribution of parasite proteins to altered mechanical properties of malaria-infected red blood cells. *Blood,* 99**,** 1060-3.
GLENISTER, F. K., FERNANDEZ, K. M., KATS, L. M., HANSSEN, E., MOHANDAS, N., COPPEL, R. L. & COOKE, B. M. 2009. Functional alteration of red blood cells by a megadalton protein of *Plasmodium falciparum*. *Blood,* 113**,** 919-28.
GUY, A. J., IRANI, V., MACRAILD, C. A., ANDERS, R. F., NORTON, R. S., BEESON, J. G., RICHARDS, J. S. & RAMSLAND, P. A. 2015. Insights into the Immunological Properties of Intrinsically Disordered Malaria Proteins Using Proteome Scale Predictions. *PLoS One,* 10**,** e0141729.
HALFMANN, R., ALBERTI, S., KRISHNAN, R., LYLE, N., O'DONNELL, C. W., KING, O. D., BERGER, B., PAPPU, R. V. & LINDQUIST, S. 2011. Opposing effects of glutamine and asparagine govern prion formation by intrinsically disordered proteins. *Mol Cell,* 43**,** 72-84.
HALFMANN, R., ALBERTI, S. & LINDQUIST, S. 2010. Prions, protein homeostasis, and phenotypic diversity. *Trends Cell Biol,* 20**,** 125-33.
HAMILTON, W. L., CLAESSENS, A., OTTO, T. D., KEKRE, M., FAIRHURST, R. M., RAYNER, J. C. & KWIATKOWSKI, D. 2016. Extreme mutation bias and high AT content in *Plasmodium falciparum*. *Nucleic Acids Res*.
HANSSEN, E., MCMILLAN, P. J. & TILLEY, L. 2010. Cellular architecture of *Plasmodium falciparum*-infected erythrocytes. *Int J Parasitol,* 40**,** 1127-35.
HAWTHORNE, P. L., TRENHOLME, K. R., SKINNER-ADAMS, T. S., SPIELMANN, T., FISCHER, K., DIXON, M. W., ORTEGA, M. R., ANDERSON, K. L., KEMP, D. J. & GARDINER, D. L. 2004. A novel *Plasmodium falciparum* ring stage protein, REX, is located in Maurer's clefts. *Mol Biochem Parasitol,* 136**,** 181-9.
HEIBER, A., KRUSE, F., PICK, C., GRURING, C., FLEMMING, S., OBERLI, A., SCHOELER, H., RETZLAFF, S., MESEN-RAMIREZ, P., HISS, J. A., KADEKOPPALA, M., HECHT, L., HOLDER, A. A., GILBERGER, T. W. & SPIELMANN, T. 2013. Identification of new PNEPs indicates a substantial non-PEXEL exportome and underpins common features in *Plasmodium falciparum* protein export. *PLoS Pathog,* 9**,** e1003546.
HERRERA, R., ANDERSON, C., KUMAR, K., MOLINA-CRUZ, A., NGUYEN, V., BURKHARDT, M., REITER, K., SHIMP, R., JR., HOWARD, R. F., SRINIVASAN, P., NOLD, M. J., RAGHEB, D., SHI, L., DECOTIIS, M., AEBIG, J., LAMBERT, L., RAUSCH, K. M., MURATOVA, O., JIN, A., REED, S. G., SINNIS, P., BARILLAS-MURY, C., DUFFY, P. E., MACDONALD, N. J. & NARUM, D. L. 2015. Reversible Conformational Change in the *Plasmodium falciparum* Circumsporozoite Protein Masks Its Adhesion Domains. *Infect Immun,* 83**,** 3771-80.
HILLER, N. L., BHATTACHARJEE, S., VAN OOIJ, C., LIOLIOS, K., HARRISON, T., LOPEZ-ESTRANO, C. & HALDAR, K. 2004. A host-targeting signal in virulence proteins reveals a secretome in malarial infection. *Science,* 306**,** 1934-7.
HODDER, A. N., MAIER, A. G., RUG, M., BROWN, M., HOMMEL, M., PANTIC, I., PUIG-DE-MORALES-MARINKOVIC, M., SMITH, B., TRIGLIA, T., BEESON, J. & COWMAN, A. F. 2009. Analysis of structure and function of the giant protein Pf332 in *Plasmodium falciparum*. *Mol Microbiol,* 71**,** 48-65.
HOFFMAN, S. L., VEKEMANS, J., RICHIE, T. L. & DUFFY, P. E. 2015. The March Toward Malaria Vaccines. *Am J Prev Med,* 49**,** S319-33.
HOUZE, S., HUBERT, V., LE PESSEC, G., LE BRAS, J. & CLAIN, J. 2011. Combined deletions of pfhrp2 and pfhrp3 genes result in *Plasmodium falciparum* malaria false-negative rapid diagnostic test. *J Clin Microbiol,* 49**,** 2694-6.
HOWARD, R. J., UNI, S., AIKAWA, M., ALEY, S. B., LEECH, J. H., LEW, A. M., WELLEMS, T. E., RENER, J. & TAYLOR, D. W. 1986. Secretion of a malarial histidine-rich protein (Pf HRP II) from *Plasmodium falciparum*-infected erythrocytes. *J Cell Biol,* 103**,** 1269-77.
HOWE, K. J. 2002. RNA polymerase II conducts a symphony of pre-mRNA processing activities. *Biochim Biophys Acta,* 1577**,** 308-24.
HUGHES, A. L. 2004. The evolution of amino acid repeat arrays in *Plasmodium* and other organisms. *J Mol Evol,* 59**,** 528-35.

JANI, D., NAGARKATTI, R., BEATTY, W., ANGEL, R., SLEBODNICK, C., ANDERSEN, J., KUMAR, S. & RATHORE, D. 2008. HDP-a novel heme detoxification protein from the malaria parasite. *PLoS Pathog,* 4**,** e1000053.

JIANG, L., MU, J., ZHANG, Q., NI, T., SRINIVASAN, P., RAYAVARA, K., YANG, W., TURNER, L., LAVSTSEN, T., THEANDER, T. G., PENG, W., WEI, G., JING, Q., WAKABAYASHI, Y., BANSAL, A., LUO, Y., RIBEIRO, J. M., SCHERF, A., ARAVIND, L., ZHU, J., ZHAO, K. & MILLER, L. H. 2013. PfSETvs methylation of histone H3K36 represses virulence genes in *Plasmodium falciparum*. *Nature,* 499**,** 223-7.

KAFSACK, B. F., ROVIRA-GRAELLS, N., CLARK, T. G., BANCELLS, C., CROWLEY, V. M., CAMPINO, S. G., WILLIAMS, A. E., DROUGHT, L. G., KWIATKOWSKI, D. P., BAKER, D. A., CORTES, A. & LLINAS, M. 2014. A transcriptional switch underlies commitment to sexual development in malaria parasites. *Nature,* 507**,** 248-52.

KAJAVA, A. V. 2012. Tandem repeats in proteins: from sequence to structure. *J Struct Biol,* 179**,** 279-88.

KATO, M., HAN, T. W., XIE, S., SHI, K., DU, X., WU, L. C., MIRZAEI, H., GOLDSMITH, E. J., LONGGOOD, J., PEI, J., GRISHIN, N. V., FRANTZ, D. E., SCHNEIDER, J. W., CHEN, S., LI, L., SAWAYA, M. R., EISENBERG, D., TYCKO, R. & MCKNIGHT, S. L. 2012. Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell,* 149**,** 753-67.

KATS, L. M., PROELLOCKS, N. I., BUCKINGHAM, D. W., BLANC, L., HALE, J., GUO, X., PEI, X., HERRMANN, S., HANSSEN, E. G., COPPEL, R. L., MOHANDAS, N., AN, X. & COOKE, B. M. 2015. Interactions between *Plasmodium falciparum* skeleton-binding protein 1 and the membrane skeleton of malaria-infected red blood cells. *Biochim Biophys Acta,* 1848**,** 1619-28.

KEMP, D. J., COPPEL, R. L. & ANDERS, R. F. 1987. Repetitive proteins and genes of malaria. *Annu Rev Microbiol,* 41**,** 181-208.

KILEJIAN, A., RASHID, M. A., AIKAWA, M., AJI, T. & YANG, Y. F. 1991. Selective association of a fragment of the knob protein with spectrin, actin and the red cell membrane. *Mol Biochem Parasitol,* 44**,** 175-81.

KISHORE, S. P., PERKINS, S. L., TEMPLETON, T. J. & DEITSCH, K. W. 2009. An unusual recent expansion of the C-terminal domain of RNA polymerase II in primate malaria parasites features a motif otherwise found only in mammalian polymerases. *J Mol Evol,* 68**,** 706-14.

KOITA, O. A., DOUMBO, O. K., OUATTARA, A., TALL, L. K., KONARE, A., DIAKITE, M., DIALLO, M., SAGARA, I., MASINDE, G. L., DOUMBO, S. N., DOLO, A., TOUNKARA, A., TRAORE, I. & KROGSTAD, D. J. 2012. False-negative rapid diagnostic tests for malaria and deletion of the histidine-rich repeat region of the hrp2 gene. *Am J Trop Med Hyg,* 86**,** 194-8.

KUMAR, N., PANDE, V., BHATT, R. M., SHAH, N. K., MISHRA, N., SRIVASTAVA, B., VALECHA, N. & ANVIKAR, A. R. 2013. Genetic deletion of HRP2 and HRP3 in Indian *Plasmodium falciparum* population and false negative malaria rapid diagnostic test. *Acta Trop,* 125**,** 119-21.

LI, J., MITAMURA, T., FOX, B. A., BZIK, D. J. & HORII, T. 2002. Differential localization of processed fragments of *Plasmodium falciparum* serine repeat antigen and further processing of its N-terminal 47 kDa fragment. *Parasitol Int,* 51**,** 343-52.

LI, P., XING, H., ZHAO, Z., YANG, Z., CAO, Y., LI, W., YAN, G., SATTABONGKOT, J., CUI, L. & FAN, Q. 2015. Genetic diversity of *Plasmodium falciparum* histidine-rich protein 2 in the China-Myanmar border area. *Acta Trop,* 152**,** 26-31.

LIAO, S. M., TAYLOR, I. C., KINGSTON, R. E. & YOUNG, R. A. 1991. RNA polymerase II carboxy-terminal domain contributes to the response to multiple acidic activators in vitro. *Genes Dev,* 5**,** 2431-40.

LINDNER, S. E., MILLER, J. L. & KAPPE, S. H. 2012. Malaria parasite pre-erythrocytic infection: preparation meets opportunity. *Cell Microbiol,* 14**,** 316-24.

LITINGTUNG, Y., LAWLER, A. M., SEBALD, S. M., LEE, E., GEARHART, J. D., WESTPHAL, H. & CORDEN, J. L. 1999. Growth retardation and neonatal lethality in mice with a

homozygous deletion in the C-terminal domain of RNA polymerase II. *Mol Gen Genet,* 261**,** 100-5.
LUNDE, B. M., REICHOW, S. L., KIM, M., SUH, H., LEEPER, T. C., YANG, F., MUTSCHLER, H., BURATOWSKI, S., MEINHART, A. & VARANI, G. 2010. Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. *Nat Struct Mol Biol,* 17**,** 1195-201.
MACRAILD, C. A., RICHARDS, J. S., ANDERS, R. F. & NORTON, R. S. 2016. Antibody Recognition of Disordered Antigens. *Structure,* 24**,** 148-57.
MAIER, A. G., RUG, M., O'NEILL, M. T., BEESON, J. G., MARTI, M., REEDER, J. & COWMAN, A. F. 2007. Skeleton-binding protein 1 functions at the parasitophorous vacuole membrane to traffic PfEMP1 to the *Plasmodium falciparum*-infected erythrocyte surface. *Blood,* 109**,** 1289-97.
MAIER, A. G., RUG, M., O'NEILL, M. T., BROWN, M., CHAKRAVORTY, S., SZESTAK, T., CHESSON, J., WU, Y., HUGHES, K., COPPEL, R. L., NEWBOLD, C., BEESON, J. G., CRAIG, A., CRABB, B. S. & COWMAN, A. F. 2008. Exported proteins required for virulence and rigidity of *Plasmodium falciparum*-infected human erythrocytes. *Cell,* 134**,** 48-61.
MALTHA, J., GAMBOA, D., BENDEZU, J., SANCHEZ, L., CNOPS, L., GILLET, P. & JACOBS, J. 2012. Rapid diagnostic tests for malaria diagnosis in the Peruvian Amazon: impact of pfhrp2 gene deletions and cross-reactions. *PLoS One,* 7**,** e43094.
MARTI, M., GOOD, R. T., RUG, M., KNUEPFER, E. & COWMAN, A. F. 2004. Targeting malaria virulence and remodeling proteins to the host erythrocyte. *Science,* 306**,** 1930-3.
MAYER, C., SLATER, L., ERAT, M. C., KONRAT, R. & VAKONAKIS, I. 2012. Structural analysis of the *Plasmodium falciparum* erythrocyte membrane protein 1 (PfEMP1) intracellular domain reveals a conserved interaction epitope. *J Biol Chem,* 287**,** 7182-9.
MCHUGH, E., BATINOVIC, S., HANSSEN, E., MCMILLAN, P. J., KENNY, S., GRIFFIN, M. D., CRAWFORD, S., TRENHOLME, K. R., GARDINER, D. L., DIXON, M. W. & TILLEY, L. 2015. A repeat sequence domain of the ring-exported protein-1 of *Plasmodium falciparum* controls export machinery architecture and virulence protein trafficking. *Mol Microbiol,* 98**,** 1101-14.
MENDES, T. A., LOBO, F. P., RODRIGUES, T. S., RODRIGUES-LUIZ, G. F., DAROCHA, W. D., FUJIWARA, R. T., TEIXEIRA, S. M. & BARTHOLOMEU, D. C. 2013. Repeat-enriched proteins are related to host cell invasion and immune evasion in parasitic protozoa. *Mol Biol Evol,* 30**,** 951-63.
MIKOLAJCZAK, S. A., SACCI, J. B., JR., DE LA VEGA, P., CAMARGO, N., VANBUSKIRK, K., KRZYCH, U., CAO, J., JACOBS-LORENA, M., COWMAN, A. F. & KAPPE, S. H. 2011. Disruption of the *Plasmodium falciparum* liver-stage antigen-1 locus causes a differentiation defect in late liver-stage parasites. *Cell Microbiol,* 13**,** 1250-60.
MILES, A., IQBAL, Z., VAUTERIN, P., PEARSON, R., CAMPINO, S., THERON, M., GOULD, K., MEAD, D., DRURY, E., O'BRIEN, J., RUANO RUBIO, V., MACINNIS, B., MWANGI, J., SAMARAKOON, U., RANFORD-CARTWRIGHT, L., FERDIG, M., HAYTON, K., SU, X. Z., WELLEMS, T., RAYNER, J., MCVEAN, G. & KWIATKOWSKI, D. 2016. Indels, structural variation, and recombination drive genomic diversity in *Plasmodium falciparum*. *Genome Res,* 26**,** 1288-99.
MURALIDHARAN, V., OKSMAN, A., IWAMOTO, M., WANDLESS, T. J. & GOLDBERG, D. E. 2011. Asparagine repeat function in a *Plasmodium falciparum* protein assessed via a regulatable fluorescent affinity tag. *Proc Natl Acad Sci U S A,* 108**,** 4411-6.
MURALIDHARAN, V., OKSMAN, A., PAL, P., LINDQUIST, S. & GOLDBERG, D. E. 2012. *Plasmodium falciparum* heat shock protein 110 stabilizes the asparagine repeat-rich parasite proteome during malarial fevers. *Nat Commun,* 3**,** 1310.
MURILLO SOLANO, C., AKINYI OKOTH, S., ABDALLAH, J. F., PAVA, Z., DORADO, E., INCARDONA, S., HUBER, C. S., MACEDO DE OLIVEIRA, A., BELL, D., UDHAYAKUMAR, V. & BARNWELL, J. W. 2015. Deletion of *Plasmodium falciparum* Histidine-Rich Protein 2 (pfhrp2) and Histidine-Rich Protein 3 (pfhrp3) Genes in Colombian Parasites. *PLoS One,* 10**,** e0131576.

NANZER, A. P., TORDA, A. E., BISANG, C., WEBER, C., ROBINSON, J. A. & VAN GUNSTEREN, W. F. 1997. Dynamical studies of peptide motifs in the *Plasmodium falciparum* circumsporozoite surface protein by restrained and unrestrained MD simulations. *J Mol Biol,* 267**,** 1012-25.

NDONWI, M., BURLINGAME, O. O., MILLER, A. S., TOLLEFSEN, D. M., BROZE, G. J., JR. & GOLDBERG, D. E. 2011. Inhibition of antithrombin by *Plasmodium falciparum* histidine-rich protein II. *Blood,* 117**,** 6347-54.

NEAFSEY, D. E., JURASKA, M., BEDFORD, T., BENKESER, D., VALIM, C., GRIGGS, A., LIEVENS, M., ABDULLA, S., ADJEI, S., AGBENYEGA, T., AGNANDJI, S. T., AIDE, P., ANDERSON, S., ANSONG, D., APONTE, J. J., ASANTE, K. P., BEJON, P., BIRKETT, A. J., BRULS, M., CONNOLLY, K. M., D'ALESSANDRO, U., DOBANO, C., GESASE, S., GREENWOOD, B., GRIMSBY, J., TINTO, H., HAMEL, M. J., HOFFMAN, I., KAMTHUNZI, P., KARIUKI, S., KREMSNER, P. G., LEACH, A., LELL, B., LENNON, N. J., LUSINGU, J., MARSH, K., MARTINSON, F., MOLEL, J. T., MOSS, E. L., NJUGUNA, P., OCKENHOUSE, C. F., OGUTU, B. R., OTIENO, W., OTIENO, L., OTIENO, K., OWUSU-AGYEI, S., PARK, D. J., PELLE, K., ROBBINS, D., RUSS, C., RYAN, E. M., SACARLAL, J., SOGOLOFF, B., SORGHO, H., TANNER, M., THEANDER, T., VALEA, I., VOLKMAN, S. K., YU, Q., LAPIERRE, D., BIRREN, B. W., GILBERT, P. B. & WIRTH, D. F. 2015. Genetic Diversity and Protective Efficacy of the RTS,S/AS01 Malaria Vaccine. *N Engl J Med,* 373**,** 2025-37.

NICOLL, W. S., SACCI, J. B., RODOLFO, C., DI GIACOMO, G., PIACENTINI, M., HOLLAND, Z. J., DOERIG, C., HOLLINGDALE, M. R. & LANAR, D. E. 2011. *Plasmodium falciparum* liver stage antigen-1 is cross-linked by tissue transglutaminase. *Malar J,* 10**,** 14.

NONET, M., SCAFE, C., SEXTON, J. & YOUNG, R. 1987. Eucaryotic RNA polymerase conditional mutant that rapidly ceases mRNA synthesis. *Mol Cell Biol,* 7**,** 1602-11.

OAKLEY, M. S., KUMAR, S., ANANTHARAMAN, V., ZHENG, H., MAHAJAN, B., HAYNES, J. D., MOCH, J. K., FAIRHURST, R., MCCUTCHAN, T. F. & ARAVIND, L. 2007. Molecular factors and biochemical pathways induced by febrile temperature in intraerythrocytic *Plasmodium falciparum* parasites. *Infect Immun,* 75**,** 2012-25.

OBERLI, A., SLATER, L. M., CUTTS, E., BRAND, F., MUNDWILER-PACHLATKO, E., RUSCH, S., MASIK, M. F., ERAT, M. C., BECK, H. P. & VAKONAKIS, I. 2014. A *Plasmodium falciparum* PHIST protein binds the virulence factor PfEMP1 and comigrates to knobs on the host cell surface. *FASEB J,* 28**,** 4420-33.

OBERLI, A., ZURBRUGG, L., RUSCH, S., BRAND, F., BUTLER, M. E., DAY, J. L., CUTTS, E. E., LAVSTSEN, T., VAKONAKIS, I. & BECK, H. P. 2016. *Plasmodium falciparum Plasmodium* helical interspersed subtelomeric proteins contribute to cytoadherence and anchor *P. falciparum* erythrocyte membrane protein 1 to the host cell cytoskeleton. *Cell Microbiol,* 18**,** 1415-28.

OLDFIELD, C. J. & DUNKER, A. K. 2014. Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu Rev Biochem,* 83**,** 553-84.

OMA, Y., KINO, Y., SASAGAWA, N. & ISHIURA, S. 2004. Intracellular localization of homopolymeric amino acid-containing proteins expressed in mammalian cells. *J Biol Chem,* 279**,** 21217-22.

PACHLATKO, E., RUSCH, S., MULLER, A., HEMPHILL, A., TILLEY, L., HANSSEN, E. & BECK, H. P. 2010. MAHRP2, an exported protein of *Plasmodium falciparum*, is an essential component of Maurer's cleft tethers. *Mol Microbiol,* 77**,** 1136-52.

PAL, P., BALABAN, A. E., DIAMOND, M. S., SINNIS, P., KLEIN, R. S. & GOLDBERG, D. E. 2017. *Plasmodium falciparum* histidine-rich protein II causes vascular leakage and exacerbates experimental cerebral malaria in mice. *PLoS One,* 12**,** e0177142.

PAL, P., DANIELS, B. P., OSKMAN, A., DIAMOND, M. S., KLEIN, R. S. & GOLDBERG, D. E. 2016. *Plasmodium falciparum* Histidine-Rich Protein II Compromises Brain Endothelial Barriers and May Promote Cerebral Malaria Pathogenesis. *MBio,* 7.

PAPALEXIS, V., SIOMOS, M. A., CAMPANALE, N., GUO, X., KOCAK, G., FOLEY, M. & TILLEY, L. 2001. Histidine-rich protein 2 of the malaria parasite, *Plasmodium falciparum*, is involved in

detoxification of the by-products of haemoglobin degradation. *Mol Biochem Parasitol,* 115**,** 77-86.
PARISH, L. A., MAI, D. W., JONES, M. L., KITSON, E. L. & RAYNER, J. C. 2013. A member of the *Plasmodium falciparum* PHIST family binds to the erythrocyte cytoskeleton component band 4.1. *Malar J,* 12**,** 160.
PATRA, A. P., SHARMA, S. & AINAVARAPU, S. R. 2017. Force Spectroscopy of the *Plasmodium falciparum* Vaccine Candidate Circumsporozoite Protein Suggests a Mechanically Pliable Repeat Region. *J Biol Chem,* 292**,** 2110-2119.
PEI, X., AN, X., GUO, X., TARNAWSKI, M., COPPEL, R. & MOHANDAS, N. 2005. Structural and functional studies of interaction between *Plasmodium falciparum* knob-associated histidine-rich protein (KAHRP) and erythrocyte spectrin. *J Biol Chem,* 280**,** 31166-71.
PERSSON, C., OLIVEIRA, G. A., SULTAN, A. A., BHANOT, P., NUSSENZWEIG, V. & NARDIN, E. 2002. Cutting edge: a new tool to evaluate human pre-erythrocytic malaria vaccines: rodent parasites bearing a hybrid *Plasmodium falciparum* circumsporozoite protein. *J Immunol,* 169**,** 6681-5.
PETERS, T. W. & HUANG, M. 2007. Protein aggregation and polyasparagine-mediated cellular toxicity in Saccharomyces cerevisiae. *Prion,* 1**,** 144-53.
PETERSEN, W., MATUSCHEWSKI, K. & INGMUNDSON, A. 2015. Trafficking of the signature protein of intra-erythrocytic *Plasmodium berghei*-induced structures, IBIS1, to *P. falciparum* Maurer's clefts. *Mol Biochem Parasitol,* 200**,** 25-9.
PLASSMEYER, M. L., REITER, K., SHIMP, R. L., JR., KOTOVA, S., SMITH, P. D., HURT, D. E., HOUSE, B., ZOU, X., ZHANG, Y., HICKMAN, M., UCHIME, O., HERRERA, R., NGUYEN, V., GLEN, J., LEBOWITZ, J., JIN, A. J., MILLER, L. H., MACDONALD, N. J., WU, Y. & NARUM, D. L. 2009. Structure of the *Plasmodium falciparum* circumsporozoite protein, a leading malaria vaccine candidate. *J Biol Chem,* 284**,** 26951-63.
PROELLOCKS, N. I., HERRMANN, S., BUCKINGHAM, D. W., HANSSEN, E., HODGES, E. K., ELSWORTH, B., MORAHAN, B. J., COPPEL, R. L. & COOKE, B. M. 2014. A lysine-rich membrane-associated PHISTb protein involved in alteration of the cytoadhesive properties of *Plasmodium falciparum*-infected red blood cells. *FASEB J,* 28**,** 3103-13.
PRZYBORSKI, J. M., NYBOER, B. & LANZER, M. 2016. Ticket to ride: export of proteins to the *Plasmodium falciparum*-infected erythrocyte. *Mol Microbiol,* 101**,** 1-11.
RAMUTTON, T., HENDRIKSEN, I. C., MWANGA-AMUMPAIRE, J., MTOVE, G., OLAOSEBIKAN, R., TSHEFU, A. K., ONYAMBOKO, M. A., KAREMA, C., MAITLAND, K., GOMES, E., GESASE, S., REYBURN, H., SILAMUT, K., CHOTIVANICH, K., PROMNARES, K., FANELLO, C. I., VON SEIDLEIN, L., DAY, N. P., WHITE, N. J., DONDORP, A. M., IMWONG, M. & WOODROW, C. J. 2012. Sequence variation does not confound the measurement of plasma PfHRP2 concentration in African children presenting with severe malaria. *Malar J,* 11**,** 276.
ROBINSON, P. J., BUSHNELL, D. A., TRNKA, M. J., BURLINGAME, A. L. & KORNBERG, R. D. 2012. Structure of the mediator head module bound to the carboxy-terminal domain of RNA polymerase II. *Proc Natl Acad Sci U S A,* 109**,** 17931-5.
RTS, S. C. T. P. 2015. Efficacy and safety of RTS,S/AS01 malaria vaccine with or without a booster dose in infants and children in Africa: final results of a phase 3, individually randomised, controlled trial. *Lancet,* 386**,** 31-45.
RUG, M., CYRKLAFF, M., MIKKONEN, A., LEMGRUBER, L., KUELZER, S., SANCHEZ, C. P., THOMPSON, J., HANSSEN, E., O'NEILL, M., LANGER, C., LANZER, M., FRISCHKNECHT, F., MAIER, A. G. & COWMAN, A. F. 2014. Export of virulence proteins by malaria-infected erythrocytes involves remodeling of host actin cytoskeleton. *Blood,* 124**,** 3459-68.
RUG, M., PRESCOTT, S. W., FERNANDEZ, K. M., COOKE, B. M. & COWMAN, A. F. 2006. The role of KAHRP domains in knob formation and cytoadherence of P falciparum-infected human erythrocytes. *Blood,* 108**,** 370-8.
SALICHS, E., LEDDA, A., MULARONI, L., ALBA, M. M. & DE LA LUNA, S. 2009. Genome-wide analysis of histidine repeats reveals their role in the localization of human proteins to the nuclear speckles compartment. *PLoS Genet,* 5**,** e1000397.

SARGEANT, T. J., MARTI, M., CALER, E., CARLTON, J. M., SIMPSON, K., SPEED, T. P. & COWMAN, A. F. 2006. Lineage-specific expansion of proteins exported to erythrocytes in malaria parasites. *Genome Biol,* 7**,** R12.

SARIDAKI, T., FROHLICH, K. S., BRAUN-BRETON, C. & LANZER, M. 2009. Export of PfSBP1 to the *Plasmodium falciparum* Maurer's clefts. *Traffic,* 10**,** 137-52.

SCHOFIELD, L. 1991. On the function of repetitive domains in protein antigens of *Plasmodium* and other eukaryotic parasites. *Parasitol Today,* 7**,** 99-105.

SCHULZE, J., KWIATKOWSKI, M., BORNER, J., SCHLUTER, H., BRUCHHAUS, I., BURMESTER, T., SPIELMANN, T. & PICK, C. 2015. The *Plasmodium falciparum* exportome contains non-canonical PEXEL/HT proteins. *Mol Microbiol,* 97**,** 301-14.

SHAKYA, B., PENN, W. D., NAKAYASU, E. S. & LACOUNT, D. J. 2017. The *Plasmodium falciparum* exported protein PF3D7_0402000 binds to erythrocyte ankyrin and band 4.1. *Mol Biochem Parasitol,* 216**,** 5-13.

SHOUBRIDGE, C., CLOOSTERMAN, D., PARKINSON-LAWERENCE, E., BROOKS, D. & GECZ, J. 2007. Molecular pathology of expanded polyalanine tract mutations in the Aristaless-related homeobox gene. *Genomics,* 90**,** 59-71.

SHOULDERS, M. D. & RAINES, R. T. 2009. Collagen structure and stability. *Annu Rev Biochem,* 78**,** 929-58.

SILVESTRINI, F., LASONDER, E., OLIVIERI, A., CAMARDA, G., VAN SCHAIJK, B., SANCHEZ, M., YOUNIS YOUNIS, S., SAUERWEIN, R. & ALANO, P. 2010. Protein export marks the early phase of gametocytogenesis of the human malaria parasite *Plasmodium falciparum*. *Mol Cell Proteomics,* 9**,** 1437-48.

SINHA, A., HUGHES, K. R., MODRZYNSKA, K. K., OTTO, T. D., PFANDER, C., DICKENS, N. J., RELIGA, A. A., BUSHELL, E., GRAHAM, A. L., CAMERON, R., KAFSACK, B. F., WILLIAMS, A. E., LLINAS, M., BERRIMAN, M., BILLKER, O. & WATERS, A. P. 2014. A cascade of DNA-binding proteins for sexual commitment and development in *Plasmodium*. *Nature,* 507**,** 253-7.

SINNIS, P. & NARDIN, E. 2002. Sporozoite antigens: biology and immunology of the circumsporozoite protein and thrombospondin-related anonymous protein. *Chem Immunol,* 80**,** 70-96.

SPIELMANN, T. & GILBERGER, T. W. 2015. Critical Steps in Protein Export of *Plasmodium falciparum* Blood Stages. *Trends Parasitol,* 31**,** 514-25.

SPILLMAN, N. J., BECK, J. R. & GOLDBERG, D. E. 2015. Protein Export into Malaria Parasite-Infected Erythrocytes: Mechanisms and Functional Consequences. *Annu Rev Biochem*.

SPYCHER, C., KLONIS, N., SPIELMANN, T., KUMP, E., STEIGER, S., TILLEY, L. & BECK, H. P. 2003. MAHRP-1, a novel *Plasmodium falciparum* histidine-rich protein, binds ferriprotoporphyrin IX and localizes to the Maurer's clefts. *J Biol Chem,* 278**,** 35373-83.

SPYCHER, C., RUG, M., PACHLATKO, E., HANSSEN, E., FERGUSON, D., COWMAN, A. F., TILLEY, L. & BECK, H. P. 2008. The Maurer's cleft protein MAHRP1 is essential for trafficking of PfEMP1 to the surface of *Plasmodium falciparum*-infected erythrocytes. *Mol Microbiol,* 68**,** 1300-14.

SULLIVAN, D. J., JR., GLUZMAN, I. Y. & GOLDBERG, D. E. 1996. *Plasmodium* hemozoin formation mediated by histidine-rich proteins. *Science,* 271**,** 219-22.

TAN, J. C., TAN, A., CHECKLEY, L., HONSA, C. M. & FERDIG, M. T. 2010. Variable numbers of tandem repeats in *Plasmodium falciparum* genes. *J Mol Evol,* 71**,** 268-78.

TARR, S. J., MOON, R. W., HARDEGE, I. & OSBORNE, A. R. 2014. A conserved domain targets exported PHISTb family proteins to the periphery of *Plasmodium* infected erythrocytes. *Mol Biochem Parasitol,* 196**,** 29-40.

THEISEN, M., SOE, S., OEUVRAY, C., THOMAS, A. W., VUUST, J., DANIELSEN, S., JEPSEN, S. & DRUILHE, P. 1998. The glutamate-rich protein (GLURP) of *Plasmodium falciparum* is a target for antibody-dependent monocyte-mediated inhibition of parasite growth in vitro. *Infect Immun,* 66**,** 11-7.

TIBURCIO, M., DIXON, M. W., LOOKER, O., YOUNIS, S. Y., TILLEY, L. & ALANO, P. 2015. Specific expression and export of the *Plasmodium falciparum* Gametocyte EXported Protein-5 marks the gametocyte ring stage. *Malar J,* 14**,** 334.

UKAEGBU, U. E. & DEITSCH, K. W. 2015. The Emerging Role for RNA Polymerase II in Regulating Virulence Gene Expression in Malaria Parasites. *PLoS Pathog,* 11**,** e1004926.

UKAEGBU, U. E., KISHORE, S. P., KWIATKOWSKI, D. L., PANDARINATH, C., DAHAN-PASTERNAK, N., DZIKOWSKI, R. & DEITSCH, K. W. 2014. Recruitment of PfSET2 by RNA polymerase II to variant antigen encoding loci contributes to antigenic variation in *P. falciparum*. *PLoS Pathog,* 10**,** e1003854.

VANNINI, A. & CRAMER, P. 2012. Conservation between the RNA polymerase I, II, and III transcription initiation machineries. *Mol Cell,* 45**,** 439-46.

VERRA, F. & HUGHES, A. L. 1999. Biased amino acid composition in repeat regions of *Plasmodium* antigens. *Mol Biol Evol,* 16**,** 627-33.

VERSTREPEN, K. J., JANSEN, A., LEWITTER, F. & FINK, G. R. 2005. Intragenic tandem repeats generate functional variability. *Nat Genet,* 37**,** 986-90.

VIGNALI, M., MCKINLAY, A., LACOUNT, D. J., CHETTIER, R., BELL, R., SAHASRABUDHE, S., HUGHES, R. E. & FIELDS, S. 2008. Interaction of an atypical *Plasmodium falciparum* ETRAMP with human apolipoproteins. *Malar J,* 7**,** 211.

WARNCKE, J. D., VAKONAKIS, I. & BECK, H. P. 2016. *Plasmodium* Helical Interspersed Subtelomeric (PHIST) Proteins, at the Center of Host Cell Remodeling. *Microbiol Mol Biol Rev,* 80**,** 905-27.

WATERKEYN, J. G., WICKHAM, M. E., DAVERN, K. M., COOKE, B. M., COPPEL, R. L., REEDER, J. C., CULVENOR, J. G., WALLER, R. F. & COWMAN, A. F. 2000. Targeted mutagenesis of *Plasmodium falciparum* erythrocyte membrane protein 3 (PfEMP3) disrupts cytoadherence of malaria-infected red blood cells. *EMBO J,* 19**,** 2813-23.

WATERMEYER, J. M., HALE, V. L., HACKETT, F., CLARE, D. K., CUTTS, E. E., VAKONAKIS, I., FLECK, R. A., BLACKMAN, M. J. & SAIBIL, H. R. 2016. A spiral scaffold underlies cytoadherent knobs in *Plasmodium falciparum*-infected erythrocytes. *Blood,* 127**,** 343-51.

WELLEMS, T. E. & HOWARD, R. J. 1986. Homologous genes encode two distinct histidine-rich proteins in a cloned isolate of *Plasmodium falciparum*. *Proc Natl Acad Sci U S A,* 83**,** 6065-9.

WENG, H., GUO, X., PAPOIN, J., WANG, J., COPPEL, R., MOHANDAS, N. & AN, X. 2014. Interaction of *Plasmodium falciparum* knob-associated histidine-rich protein (KAHRP) with erythrocyte ankyrin R is required for its attachment to the erythrocyte membrane. *Biochim Biophys Acta,* 1838**,** 185-92.

WEST, M. L. & CORDEN, J. L. 1995. Construction and analysis of yeast RNA polymerase II CTD deletion and substitution mutations. *Genetics,* 140**,** 1223-33.

WHO 2016. World Malaria Report.

WICKERT, H. & KROHNE, G. 2007. The complex morphology of Maurer's clefts: from discovery to three-dimensional reconstructions. *Trends Parasitol,* 23**,** 502-9.

WURTZ, N., FALL, B., BUI, K., PASCUAL, A., FALL, M., CAMARA, C., DIATTA, B., FALL, K. B., MBAYE, P. S., DIEME, Y., BERCION, R., WADE, B., BRIOLANT, S. & PRADINES, B. 2013. Pfhrp2 and pfhrp3 polymorphisms in *Plasmodium falciparum* isolates from Dakar, Senegal: impact on rapid malaria diagnostic tests. *Malar J,* 12**,** 34.

YAGI, M., BANG, G., TOUGAN, T., PALACPAC, N. M., ARISUE, N., AOSHI, T., MATSUMOTO, Y., ISHII, K. J., EGWANG, T. G., DRUILHE, P. & HORII, T. 2014. Protective epitopes of the *Plasmodium falciparum* SERA5 malaria vaccine reside in intrinsically unstructured N-terminal repetitive sequences. *PLoS One,* 9**,** e98460.

YOH, S. M., LUCAS, J. S. & JONES, K. A. 2008. The Iws1:Spt6:CTD complex controls cotranscriptional mRNA biosynthesis and HYPB/Setd2-mediated histone H3K36 methylation. *Genes Dev,* 22**,** 3422-34.

ZAVALA, F., TAM, J. P., HOLLINGDALE, M. R., COCHRANE, A. H., QUAKYI, I., NUSSENZWEIG, R. S. & NUSSENZWEIG, V. 1985. Rationale for development of a synthetic vaccine against *Plasmodium falciparum* malaria. *Science,* 228**,** 1436-40.

ZEHRING, W. A., LEE, J. M., WEEKS, J. R., JOKERST, R. S. & GREENLEAF, A. L. 1988. The C-terminal repeat domain of RNA polymerase II largest subunit is essential in vivo but is not required for accurate transcription initiation in vitro. *Proc Natl Acad Sci U S A,* 85**,** 3698-702.

ZILVERSMIT, M. M., VOLKMAN, S. K., DEPRISTO, M. A., WIRTH, D. F., AWADALLA, P. & HARTL, D. L. 2010. Low-complexity regions in *Plasmodium falciparum*: missing links in the evolution of an extreme genome. *Mol Biol Evol,* 27**,** 2198-209.

Figure legends:

Figure 1. Repeat expansion and contraction during DNA replication and unequal crossing over in meiosis.

Figure 2. The *Plasmodium falciparum* life cycle. A) Sporozoites are injected from the salivary glands of the *Anopheles* mosquito and migrate to the host liver. B) Upon invasion of hepatocytes, thousands of merozoites are produced and subsequently released into the bloodstream. C) Merozoites invade host erythrocytes and cycle through the ring, trophozoite and schizont stages before daughter merozoites egress from the host erythrocyte and invade new cells. D) A small proportion of parasites differentiate into male and female gametocytes. E) Gametocytes are ingested into the mosquito's midgut as it takes a blood meal where they form gametes and fuse to form a diploid zygote. The zygote differentiates into an ookinete which develops into an oocyst within the midgut wall. Upon oocyst rupture, sporozoites are released into the hemolymph and travel to the mosquito salivary gland. Proteins containing repetitive/low complexity sequences are labelled at the approximate life stage where they are utilised.

Figure 3. RNA Polymerase II C-terminal domain – phosphorylation and recruitment of proteins by the RNA pol II CTD during transcription

Figure 4. Repetitive proteins exported into the host cell. Many proteins are exported into the host erythrocyte by *P. falciparum*. These are involved in creating new structures within the host cell: membranous structures called Maurer's clefts (MC) appear in the host cell cytoplasm, a tubovesicular network' (TVN) extends from the parasitophorous vacuole (PV), and 'knob' protrusions form on the cell surface which present the cytoadhesive surface protein PfEMP1. Many repetitive proteins interact with the spectrin cytoskeleton of the erythrocyte resulting in increased cell rigidity.

Figure 5. Expansion and contraction of lysine-rich repetitive sequences and changes in protein localization or protein binding affinity

The top panel illustrates the correlation between the length of a lysine-rich sequence from the protein GARP and the efficiency with which it is targeted to the erythrocyte periphery.

The lower panel illustrates the changes in binding affinity between β-spectrin and KAHRP repeats as the number of repeats in KAHRP is altered.

Gene Identifier	Alias	Consensus Motif in 3D7	Number of Repeats	Localisation
PF3D7_0930300	MSP1	GASAQS (GSGGSVASG/ SGGSVT)*	2-6 (3-5/ 4)	Merozoite surface
PF3D7_0206800	MSP2	AGGS (TTTESNSPSPPI PAGAGASGNP/AGAGASGS/ GASGSG/SGSAGG/ GAGAS/SGSAG) *	7-11 (4 6/8 13/6 9/3)	Merozoite surface
PF3D7_0207600	SERA5	QGSTGASP	3-12	PV + Merozoite surface
PF3D7_1035300	GLURP	EILPEDKNEKVQHEIVEVE SEKSVSEPAEHVEIV	4-14 3	Merozoite surface
PF3D7_1036400	LSA1	DLEQERLAKEKLQEQQS	32-93	Liver Stage
PF3D7_0318200	RNA polymerase 2	YSPTSPK	6-11	Nucleus
PF3D7_0831800	HRP2	AHHAAD	38-46	RBC cytoplasm
PF3D7_1222600	AP2-G **	KNN	7-9	Nucleus
		DTYN	16-18	
PF3D7_1370300	MAHRP1	HDHD	3-5	Maurer's Clefts
		DHG	6-12	
PF3D7_0304600	CSP	NANP	40-46	Sporozoite Surface
PF3D7_0935900	REX1	KPQAEKDASKLTTTYDQTKEV	3-6	Maurer's Clefts
		NKETKPQNDKYTL	2	
PF3D7_0501300	SBP1	ASGIGNLVGDA	5-6	Maurer's Clefts
		QNAQ	7-14	
PF3D7_1149000	Pf332	PVEEKNVSEEI	6-11	Maurer's Clefts
		FVTGELPEEDIINEKVQEEEE	3-5	
		EESASEEIVEDEGSV	5-7	

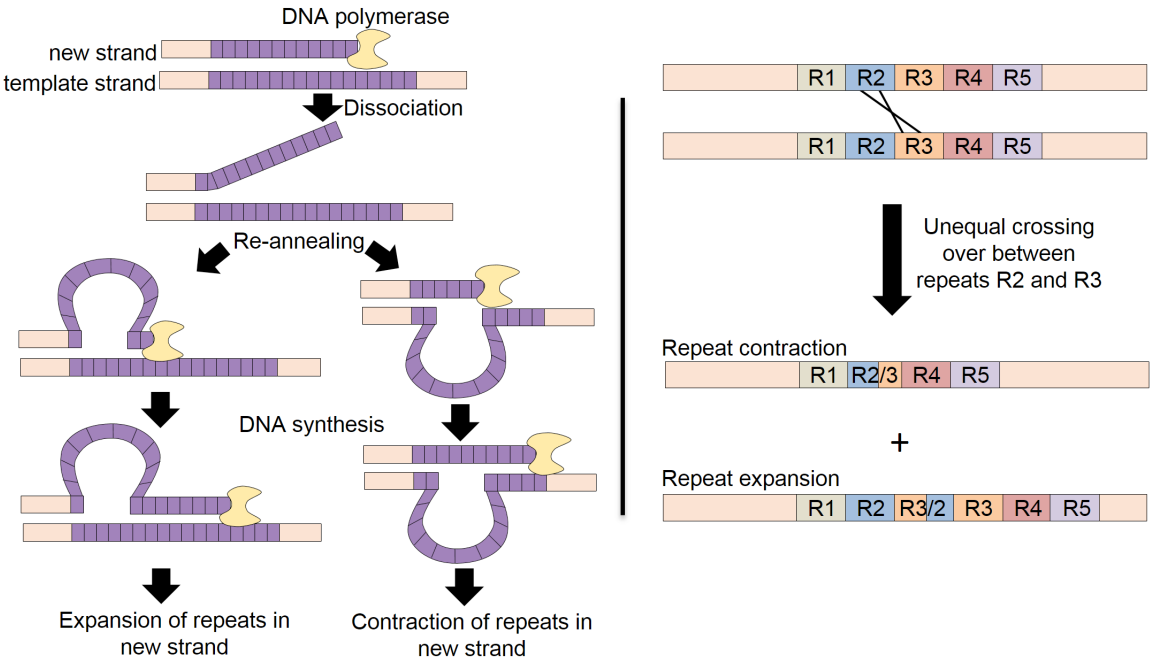
		ENVEEKKTMDEEIVDQGSVV	3-4	
		TEEVVEEEGSV	5-10	
		EEVVEEGSAT	2-11	
		EEIVEEEESSS	11-12	
		EEIVEEEGSVV	10-24	
		SVTEELVDEG	2-3	
		TEEIVED-EGSF	6-7	
		NEEILEEEGSY	6-11	
		GSATDYFVGQGSDNEEIIEE	2-4	
		IKEEQLDSEE	6-20	
		EVEEVSVDD	2-5	
		EEIEEIESVT	2-4	
		TEDVEEVSS	4-8	
PF3D7_0401800	PFD80	STA	9-18	Maurer's Clefts
		RSASAASTT	8-12	
		STSTTQSPST	3-6	
PF3D7_0102200	RESA	EEPTVADEHV	6-8	RBC periphery
		EENV	41-47	
PF3D7_0113000	GARP	EKK	12-17	RBC periphery
		EKEKKKQ	7	
		EEHKE	8-9	
		KGKKD	5	
		EEDEDDA	8-10	
PF3D7_0201900	PfEMP3	LEEYNETDLAKGKEVTNKAHEN	17-19	RBC periphery
		KNKELQNKGSEGLKENAEL	9-12	
		NKDISNKDMKNKELL	2-3	
		QQNTGLKNTPSEG	54-87	
PF3D7_0202000	KAHRP	SKKHKDHDGEKKK	4	RBC periphery
		ATKEASTSKE	4-7	
PF3D7_0402000	PHISTa	KQGGKKEEV	9-14	RBC periphery

PF3D7_0500800	MESA	GESKET	11-22	RBC periphery
		EKNDEKKDKVLGEGDKEDVK	4	
		KEKEEV	7-11	
		KEKEEV	3-9	
		ESEE	17-26	
PF3D7_0532400	LYMP	NKKVRGA	5	RBC periphery
		ENKKAGT	5-7	
PF3D7_1102300	N/A	ERKEREEREKK	9	RBC periphery
		EREKREKKEKE	13-14	
PF3D7_1148700	GEXP12	KECVPNECMK	8	RBC periphery
PF3D7_1201000	PHISTb/c	EKDEK	18-36	RBC periphery
		DDDDEDDED	7-8	
PF3D7_1476200	PHISTb	KEQEKEKERKRKE	4	RBC periphery
PF3D7_1301400	HYP12	KKKEKQE	8	RBC cytoplasm
		NEDE	12	

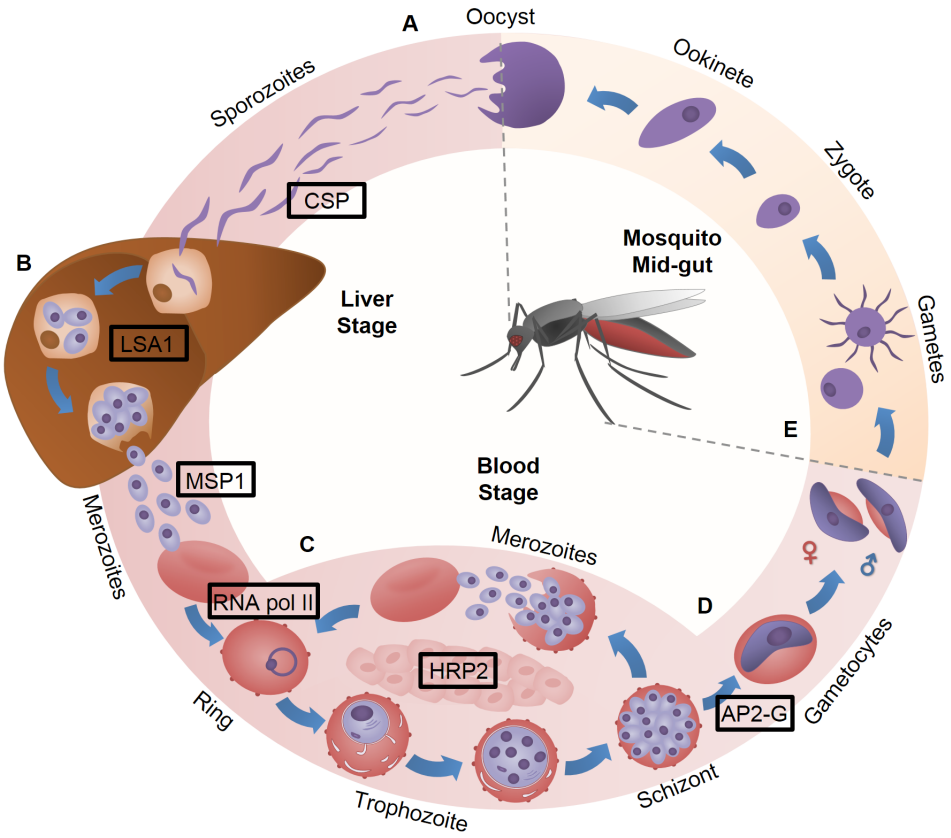
Table 1. Repetitive sequences in *Plasmodium falciparum* proteins
The repeat consensus for each protein was determined from *P. falciparum* 3D7 strain sequences using the programme XSTREAM, with gaps removed for clarity. The phases of some repeats were adjusted to those described in the literature. (*) Where the repeating motif differs between isolates the alternative repeating motifs are represented after the 3D7 consensus sequence. Variation in repeat numbers seen in 16 different *P. falciparum* strains (3D7, IT, HB3, DD2, 7G8, CD1, GA1, GB4, GN1, KE1, KH1, KH2, ML1, SD1, SN1, TG1, where two-letter codes represent the country of origin for the 11 field isolates) were determined manually using Jalview with TCOFFEE used to align the sequences. (**) AP2-G also contains several poly-N sequences which vary by up to 12 residues. Long-read PACBIO genome sequencing data was obtained from the Pf3K consortium.

A. Strand slippage during DNA replication

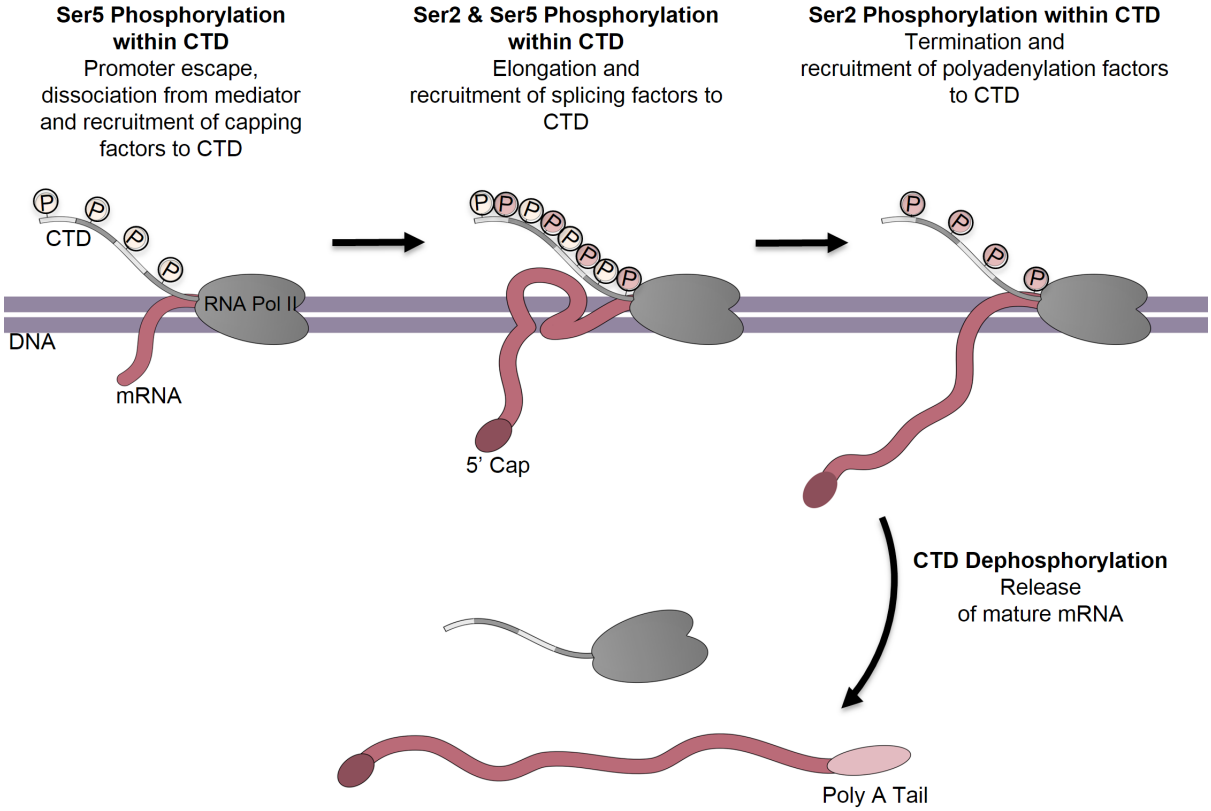
B. Unequal crossing over between homologous chromosomes during meiosis



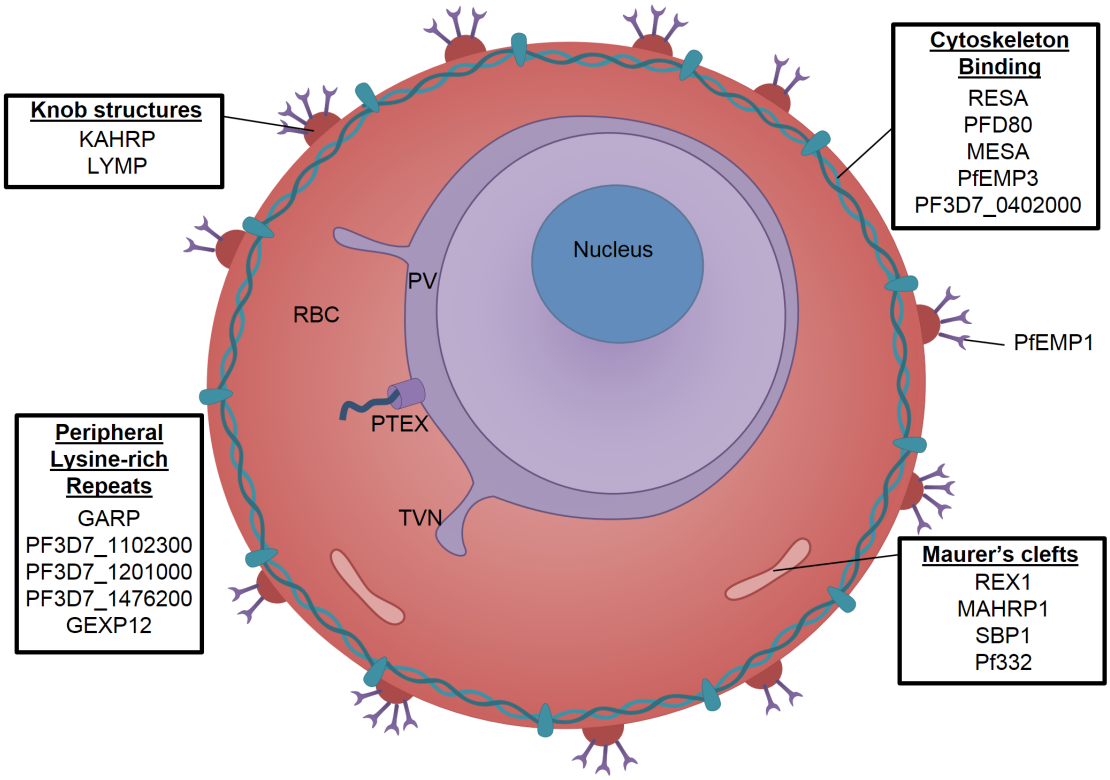
Davies et al. Figure 1



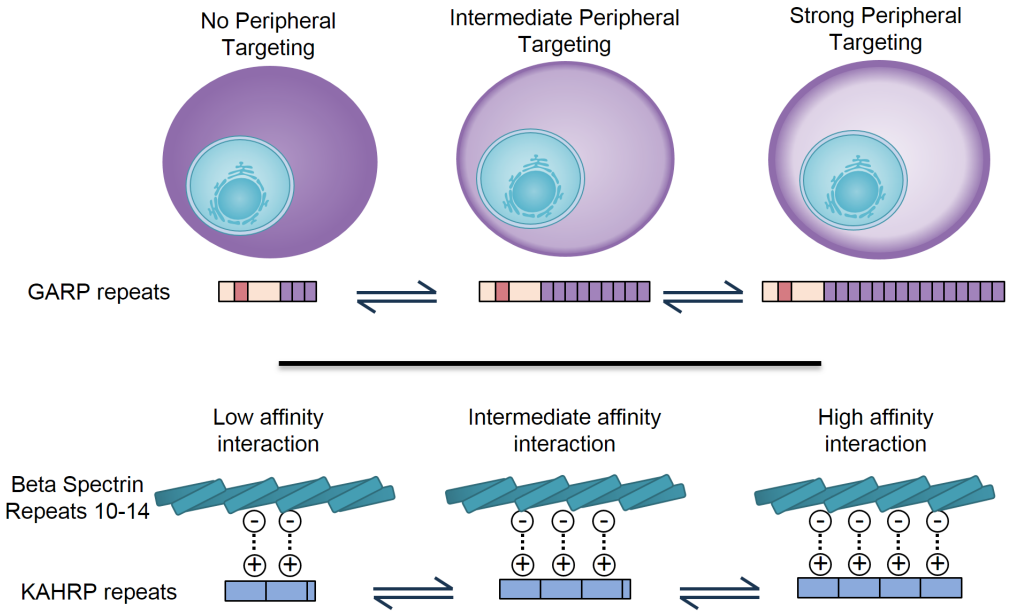
Davies et al. Figure 2



Davies et al. Figure 3



Davies et al. Figure 4



Davies et al. Figure 5